377/1.9.70

Psychometrika

VOLUME XXIV—1959
JANUARY-DECEMBER

Editorial Council

Chairman:—Harold Gulliksen

Editors:—Dorothy Adkins Wood
Paul Horst

Managing Editor:—
Lyle V. Jones

Assistant Managing Editor:—
B. J. Winer

Editorial Board

R. L. Anderson
T. W. Anderson
Rolf Bargmann
J. B. Carroll
H. S. Conrad
C. H. Coombs
L. J. Cronbach
E. E. Cureton
Paul S. Dwyer
Allen Edwards
Max D. Engelhart
Wm. K. Estes
Henry E. Garrett
Bert F. Green
J. P. Guilford
Harold Gulliksen
Paul Horst
Lloyd G. Humphreys
Truman L. Kelley
Albert K. Kurtz
Frederic M. Lord
Irving Lorge
Quinn McNemar
George A. Miller
Wm. G. Mollenkopf
Lincoln E. Moses
Frederick Mosteller
George E. Nicholson
M. W. Richardson
R. L. Thorndike
Warren S. Torgerson
Ledyard Tucker
D. F. Votaw, Jr.
Dorothy Adkins Wood

PUBLISHED QUARTERLY
By THE PSYCHOMETRIC SOCIETY
AT 1407 SHERWOOD AVENUE
RICHMOND 5, VIRGINIA

EAU OF EDNL. & PSYL. RESE

1960 INTERNATIONAL CONGRESS FOR LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

An International Congress for Logic, Methodology, and Philosophy of Science will be held at Stanford University from August 24 to September 2, 1960, under the auspices of the *International Union for History and Philosophy of Science*. The proceedings of the Congress will be organized into the following sections.

1. Mathematical logic
2. Foundations of mathematical theories
3. Philosophy of logic and mathematics
4. General problems of methodology and philosophy of science
5. Foundations of probability and induction
6. Methodology and philosophy of physical sciences
7. Methodology and philosophy of biological and psychological sciences
8. Methodology and philosophy of social sciences
9. Methodology and philosophy of linguistics
10. Methodology and philosophy of historical sciences
11. History of logic, methodology, and philosophy of science

The proceedings will consist of a number of invited addresses, in addition to brief contributed papers. The closing date for submission of abstracts of contributed papers is March 1, 1960. Information about membership fees and other details of the Congress may be obtained by writing Professor Patrick Suppes, Serra House, Stanford, California.

7.9.70
J – 377

Psychometrika

CONTENTS

STATISTICAL INFERENCES ABOUT TRUE SCORES. 1
Frederic M. Lord

A STATISTICAL TEST FOR THE SIGNIFICANCE OF A COEFFICIENT OF REPRODUCIBILITY 19
Philip C. Sagi

SIMPLE STATISTICAL METHODS FOR SCALOGRAM ANALYSIS 29
Leo A. Goodman

HIERARCHICAL FACTOR SOLUTIONS WITHOUT ROTATION 45
Robert J. Wherry

A GENERALIZATION OF THE BUSH-MOSTELLER MODEL WITH SOME SIGNIFICANCE TESTS. 53
Mary I. Hanania

SOME IMPLICATIONS OF THE LOGICAL CALCULUS FOR EMPIRICAL CLASSES FOR SOCIAL SCIENCE METHODOLOGY. 69
William C. Shutz

NOTE ON $\phi/\phi_{\max}$ 89
Edward E. Cureton

BOOK REVIEW

Clyde H. Coombs and Richard C. Kao. *Nonmetric Factor Analysis* 93
Review by Samuel Messick

VOLUME TWENTY-FOUR MARCH 1959 NUMBER 1

To establish a complete permanent set of *Psychometrika* for the Office of the Managing Editor, there remains a need for one copy of Volume 8, Number 3, September, 1943. Any reader able to provide a copy of this issue for sale to the Psychometric Corporation is requested to write to the Managing Editor, Lyle V. Jones, Box 547, Chapel Hill, North Carolina.

The Institute of Mathematical Statistics and the University of Chicago have established a series of publications entitled *Statistical Research Monographs.*

The primary purpose of this series is to provide a medium of publication for material of interest to statisticians that is not ordinarily provided for by existing media. Among the kinds of publications envisaged are:

new research results too lengthy for the usual journal article;
research results of interest in both theoretical and applied statistics;
expository monographs in particular areas of statistics;
discussions of statistical problems and techniques in particular areas of application.

Authors are invited to send manuscripts and correspondence concerning the series to Leo A. Goodman, Department of Statistics, University of Chicago, Chicago 37, Illinois.

Psychometrika

CONTENTS

ON METHODS IN THE ANALYSIS OF PROFILE DATA . . . 95
SAMUEL W. GREENHOUSE AND SEYMOUR GEISSER

DOMAIN SAMPLING FORMULATION OF CLUSTER AND FACTOR ANALYSIS 113
ROBERT C. TRYON

LEAST SQUARES ESTIMATION IN FINITE MARKOV PROCESSES 137
ALBERT MADANSKY

AN AUGMENTED MODEL FOR SPONTANEOUS REGRESSION AND RECOVERY 145
DAVID McCONNELL

A MODEL FOR ORDERED METRIC SCALING BY COMPARISON OF INTERVALS 157
ROBERT F. FAGOT

A NOTE ON FACTOR ANALYSIS: ARBITRARY ORTHOGONAL TRANSFORMATIONS 169
EDWARD E. CURETON

RANDOMLY PARALLEL TESTS AND LYERLY'S ASSUMPTION FOR THE KUDER-RICHARDSON FORMULA (21) 175
FREDERIC M. LORD

ESTIMATING ITEM INDICES BY NOMOGRAPHS 179
ROBERT M. COLVER

BOOK REVIEW

OLÉRON, PIERRE. *Les Composantes de l'Intelligence d'après les Recherches Factorielles.* 187
Review by R. DARRELL BOCK

VOLUME TWENTY-FOUR **JUNE 1959** **NUMBER 2**

Psychometrika

CONTENTS

ITEM SELECTION PROCEDURES FOR ITEM VARIABLES WITH A KNOWN FACTOR STRUCTURE 189
G. Elfving, R. Sitgreaves, and H. Solomon

A METRIC AND AN ORDERING ON SETS 207
Frank Restle

MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS 221
A. E. Maxwell

THREE MULTIVARIATE MODELS: FACTOR ANALYSIS, LATENT STRUCTURE ANALYSIS, AND LATENT PROFILE ANALYSIS 229
W. A. Gibson

STRATEGIES AND LEARNING MODELS 253
Steven J. Bryant and John C. Marica

A NOTE ON TRYON'S MEASURE OF RELIABILITY 257
Allen L. Edwards

AN IMPROVED PROCEDURE FOR THE WHERRY-WINER METHOD FOR FACTORING LARGE NUMBERS OF ITEMS 261
Leroy Wolins

GENERATING VARIABLES WITH ARBITRARY PROPERTIES 265
Paul J. Hoffman

A NOTE ON THE TRYON-KAISER SOLUTION FOR THE COMMUNALITIES 269
Henry F. Kaiser

A NOTE ON THE USE OF TRIADS FOR PAIRED COMPARISONS 273
R. E. Schucker

Continued on next page

VOLUME TWENTY-FOUR SEPTEMBER 1959 NUMBER 3

Psychometrika

CONTENTS (*Cont.*)

BOOK REVIEWS

DONALD DAVIDSON, PATRICK SUPPES, AND SIDNEY SIEGEL. *Decision Making, An Experimental Approach* 277
Review by R. M. THRALL

WARREN S. TORGERSON. *Theory and Methods of Scaling* 277
Review by WENDELL R. GARNER

D. A. S. FRASER. *Nonparametric Methods in Statistics* 279
Review by I. RICHARD SAVAGE

PHILIP J. MCCARTHY. *Introduction to Statistical Reasoning* . . . 280
Review by LESLIE KISH

VOLUME TWENTY-FOUR SEPTEMBER 1959 NUMBER 3

Psychometrika

CONTENTS

AN APPROACH TO MENTAL TEST THEORY 283
Frederic M. Lord

THEORETICAL MODELS OF CHOICE AND STRATEGY BEHAVIOR: STABLE STATE BEHAVIOR IN THE TWO-CHOICE UNCERTAIN OUTCOME SITUATION . . . 303
Sidney Siegel

THE CONCEPT OF CONFIGURATION OF INTERPERSONAL RELATIONS IN A GROUP AS A TIME-DEPENDENT STOCHASTIC PROCESS 317
Leo Katz and Charles H. Proctor

PATTERN VARIANTS ON A SQUARE FIELD 329
S. J. Prokhovnik

A SYNTHESIS OF TWO FACTOR ANALYSES OF INTERMEDIATE ALGEBRA 343
William E. Kline

REDUNDANCY IN TASK ASSIGNMENTS AND GROUP PERFORMANCE 361
Robert B. Zajonc and William H. Smoke

A NOTE ON THE USE OF MOOD'S LIKELIHOOD RATIO TEST FOR ITEM ANALYSES INVOLVING 2 × 2 TABLES WITH SMALL SAMPLES 371
Irving B. Weiner

PUBLICATIONS RECEIVED FOR REVIEW 373

BOOK REVIEWS

Jane Loevinger. *Objective Tests as Instruments of Psychological Theory* 375
Review by Robert M. W. Travers

L'Analyse Factorielle et ses Applications 376
Review by Philip H. DuBois

(cont.)

VOLUME TWENTY-FOUR DECEMBER 1959 NUMBER 4

Psychometrika

CONTENTS (*Cont.*)

MEMBERS OF THE PSYCHOMETRIC SOCIETY 379

MINUTES OF THE 1959 ANNUAL BUSINESS MEETING OF THE PSYCHOMETRIC SOCIETY 399

TREASURER'S REPORT, PSYCHOMETRIC SOCIETY 402

TREASURER'S REPORT, PSYCHOMETRIC CORPORATION . 403

INDEX FOR VOLUME 24 405

VOLUME TWENTY-FOUR DECEMBER 1959 NUMBER 4

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

STATISTICAL INFERENCES ABOUT TRUE SCORES*

FREDERIC M. LORD

EDUCATIONAL TESTING SERVICE

Formulas are derived for unbiased sample estimators of any raw or central moment of the frequency distribution of true scores. A general method is developed for obtaining from each examinee's observed score a least squares estimate of his true score.

The ordinary "observed" test score frequently can be used in practice without thought of separating it into what must logically be its component parts: true score and error of measurement. From a scientific point of view, however, the entire concern must be with true score; the observed score can be of scientific interest only as it leads to inferences and generalizations regarding true scores. Some new methods for making such generalizations are reported in the present article under two separate headings: "Inferring the Shape of the Frequency Distribution of True Scores" and "Estimating the True Score of Each Examinee." The methods given here for inferring the true-score moments are based on different mathematical models than that used in a previous derivation [10]; the formulas presented here are considerably easier to apply in actual practice.

The matrix-sampling model, used in all but the first two sections of this article, readily yields a wide variety of basic results of importance to test theory. The shape of the bivariate frequency distribution of *observed* scores on two parallel forms of the same test, for example, can be predicted from the data provided by either form alone. Again, the relationship between true scores on two different tests can be estimated from observed data; in particular, the question can be considered whether the true scores have a perfect *curvilinear* relationship. Although it is easy to obtain estimators for these and other purposes, it is more difficult to obtain usable formulas for their sampling errors, at least in the case of those estimators that involve terms above the second degree. (Some standard error formulas for second-degree estimators are given in [9].) Since a knowledge of the sampling errors of the estimators is important, it is hoped to obtain formulas for more of these before further estimation equations are published.

In the present article, formulas are derived (see Rule 1) for unbiased sample estimators of any raw or central moment of the frequency distribution of true scores. Explicit and relatively convenient computational formulas are given for both raw and central moments up to and including the fourth order. Methods for fitting frequency curves to a set of estimated true-score moments are briefly discussed.

---

*This research was carried out under contract Nonr-2214(00) with the Office of Naval Research, Department of the Navy.

A general method is developed for obtaining from each examinee's observed score a least squares estimate of his true score. Detailed computational formulas for making such estimates are given for the case where the regression of true score on observed score is assumed to be linear and for the case where this regression is assumed to be a third-degree parabola. The method can be extended to the case where this regression may be approximated by a parabola of any specified degree.

Inferring the Shape of the Frequency Distribution of True Scores

The true score of an individual is customarily estimated by the linear regression equation ([7], eq. 11:20)

$$\xi_a = \bar{x} + r_{xx}(x_a - \bar{x}), \tag{1}$$

where ξ_a is the estimated true score of examinee a, r_{xx} is the test reliability, x_a is the observed score of examinee a, and $\bar{x}$ is the mean score of the group of examinees to which examinee a belongs. Since the relationship between ξ_a and x_a in (1) is linear, the frequency distribution of ξ_a is the same as that of x_a except for origin and unit of measurement. Equation (1), therefore, does not provide a useful method for estimating the shape of the distribution of true scores.

The Assumption of Normally Distributed Errors

A first approximation to estimating the true-score distribution can be obtained by assuming the errors of measurement, e, to be distributed normally, independently of true score, with a mean of zero and a standard deviation, σ_e, which may be estimated experimentally. In this oversimplified situation, a very neat relationship (first called to the writer's attention by Professor Max Woodbury) exists between the true-score distribution and the observed-score distribution.

Let $\phi_\xi(t)$ be the characteristic function ([8], vol. 1, ch. 4) of the true-score distribution. Since e is normally distributed, its c.f. is $\exp\left(-\frac{1}{2}t^2\sigma_e^2\right)$. By definition, $x_a = e_a + \xi_a$. A well-known theorem states that the c.f. of a sum of two independently distributed variables is equal to the product of their c.f.'s. The c.f. of x is thus

$$\phi_x(t) = \phi_\xi(t) \exp\left(-\tfrac{1}{2}t^2\sigma_e^2\right), \tag{2}$$

and the cumulant generating function for x is

$$\begin{aligned} \psi_x(t) &= \log \phi_x(t) \\ &= \psi_\xi(t) - \tfrac{1}{2}t^2\sigma_e^2 . \end{aligned} \tag{3}$$

Since the rth cumulant of the distribution of x is the coefficient of $(it)^r/r!$ in the Taylor expansion of $\psi_x(t)$, it is clear that the true-score distribution and the observed-score distribution have identical cumulants except for the second; the variance (second cumulant) of the true-score distribution is equal to the variance of the observed-score distribution minus the variance of the errors of measurement, as is, of course, well known from other lines of reasoning.

Given the moments of the observed-score distribution and an experimentally determined estimate of the standard error of measurement, estimates of all the moments of the true-score distribution can thus be readily written down from standard formulas relating moments to cumulants ([8], eqs. 3.30 and 3.34). (The shape of a distribution is ordinarily determined by its moments. Methods for doing this will be discussed briefly at a later point.)

The foregoing convenient method for determining the shape of the true-score distribution may be a good approximation in the usual situation where the observed-score distribution is nearly normal. However, in this case the problem has little special interest, since the true-score distribution will then also be approximately normal with known mean and variance.

Actually, it is clear that when an examinee's true score is close to zero or to n, the number of items in the test, the error of measurement in his observed score is likely to have a distribution that is skewed and a variance smaller than that for an examinee with a less extreme true score. Some more satisfactory assumption about the nature and frequency distribution of the errors of measurement is needed.

The Matrix-Sampling Model

The matrix-sampling model ([9], sec. 8) completely abandons the assumption that the errors of measurement are distributed normally and independently of true score. Instead, the group of examinees tested is considered as drawn at random from a very large population of examinees ("type-1" sampling) and the test is considered as a sample of items drawn at random from a very large pool of items ("type-2" sampling). The error of measurement in each test score is, by definition, the sampling fluctuation arising from the type-2 sampling process; its frequency distribution and other properties are then inferred by means of standard sampling theory. Methods for estimating true scores are then readily devised.

It is assumed that the score on the test is either the number of items answered correctly, i.e., $x_a = \sum_{i=1}^{n} x_{ia}$, where $x_{ia} = 1$ if examinee a answers item i correctly, and $x_{ia} = 0$ otherwise; or, more conveniently, the proportion of items answered correctly, to be denoted by

$$z_a = \frac{1}{n} \sum_{i=1}^{n} x_{ia} = \frac{x_a}{n}. \tag{4}$$

Unless otherwise specified, the "score" of a test will hereafter refer to the *proportion-correct* rather than to the *number-correct*.

For convenience, this notation suppresses a third subscript, u, which must always be understood: x_{uia} is the score of examinee a on the ith item in test u. Note that the subscript i identifies a unique item only when a specified test (sample of items) is under consideration; otherwise, the subscript i merely denotes the ith position in any test u. In general there will be only one test for which real data are actually available; all other tests are hypothetical samples of items drawn at random from a hypothesized item pool.

The symbol E_2 will denote the operation of taking the *expected value* of a statistic in type-2 sampling, i.e., the operation of taking the average value of the statistic over all possible samples of n items each. For example, $E_2 x_{ia}$ is the equivalent of

$$\lim_{U=\infty} \frac{1}{U} \sum_{u=1}^{U} x_{uia} .$$

Thus the symbol E can be manipulated in the same way as an ordinary summation sign. (Note in particular that the expected value of a sum of variables is equal to the sum of the expected values of the variables, and that the expected value of a product of *uncorrelated* variables is equal to the product of the expected values of the variables.)

The expected value over all examinees in the population will be denoted by E_1 . The expected value taken simultaneously over all possible combinations of n-item tests and over all examinees in the population will be denoted by E_{12} . The average value over all examinees in the group (sample of examinees) tested will be denoted by A_a (or A_b) when the number, N, of examinees is very large and by the equivalent operator $N^{-1} \sum_{a=1}^{N}$, otherwise. When N is very large, the effect of the operators E_1 and A_a is identical, but it seems wise to use different symbols for the two different operations.

The first problem will be to obtain estimates of the moments of the true-score distribution that are unbiased in type-2 sampling. For convenience, these will be called *type-2-unbiased* estimates of the moments of the true scores. (By definition, an unbiased estimate is one whose expected value is equal to the parameter being estimated.) The formulas for all of these unbiased estimates, together with the moments of their sampling distributions, can be conveniently obtained by methods developed by Tukey [15, 16], Hooke [5, 6], and Robson [13]. The present derivations, however, will be carried through by means of procedures and terminology more familiar to the average reader.

Some further notation may be defined at this point.

True proportion-score of examinee a (his mean proportion-score over all possible tests):

$$\zeta_a = E_2 z_a = E_2 x_{ia} \, . \tag{5}$$

Difficulty of item i (proportion of examinees answering item i correctly), for the population of examinees and for the sample, respectively:

$$\pi_i = E_1 x_{ia} \, , \qquad p_i = A_a x_{ia} = \frac{1}{N} \sum_{a=1}^{N} x_{ia} \, . \tag{6}$$

Average true proportion-score = average item difficulty:

$$\bar{\zeta} = \bar{\pi} = E_1 \zeta_a = E_2 \pi_i = E_1 E_2 z_a = E_1 E_2 x_{ia} = E_{12} x_{ia} \, . \tag{7}$$

Proportion of examinees answering correctly all of the r items $i_1, i_2, \cdots, i_r$:

$$\begin{aligned} \pi_{i_1 i_2 \cdots i_r} &= E_1 x_{i_1 a} x_{i_2 a} \cdots x_{i_r a} \, , \\ p_{i_1 i_2 \cdots i_r} &= A_a x_{i_1 a} x_{i_2 a} \cdots x_{i_r a} \, . \end{aligned} \tag{8}$$

g, h, i, j will be used as alternative subscripts to designate items; a and b will be used as subscripts to designate examinees. The symbol $n^{[r]}$ will denote the product $n(n-1)(n-2) \cdots (n-r+1)$.

Type-2-Unbiased Estimates of Raw True-Score Moments. Now consider, for example, the quantity

$$\frac{1}{n^{[3]}} \overset{n(n-1)(n-2)}{\sum \sum_{g \neq i \neq j} \sum} \pi_g \pi_{ij} \, .$$

(The product above the summation signs indicates the number of terms in the summation.) The expected value of this quantity in type-2 sampling is

$$\begin{aligned} E_2 \frac{1}{n^{[3]}} \sum \sum_{g \neq i \neq j} \sum \pi_g \pi_{ij} &= \frac{1}{n^{[3]}} \sum \sum_{g \neq i \neq j} \sum E_2 (E_1 x_{ga})(E_1 x_{ib} x_{jb}) \\ &= E_1 E_2 x_{ga} x_{ib} x_{jb} \\ &= E_1 \zeta_a \zeta_b^2 \\ &= (E_1 \zeta_a)(E_1 \zeta_b^2) \\ &= \mu_1' \mu_2' \, . \end{aligned} \tag{9}$$

The steps in the foregoing reduction are as follows. (i) The E_2 is moved to the right and the definitions of π_g and π_{ij} are substituted for these two symbols. (ii) The value following the summation signs is the same for any g, i, and j; since there are $n(n-1)(n-2)$ terms under the summation signs, everything preceding E_2 cancels. The E_1 symbols are moved to the left (the population of examinees is assumed to be so large that terms with $a = b$ may be ignored). (iii) If a person's score on the ith item on a particular test (random sample of items) is higher than on other tests, this is no reason why his score (or another examinee's score) on the jth item should be higher than on other tests; hence x_{ib} and x_{jb} (and also x_{ga}) fluctuate independently in type-2 sampling, and the expectation of their product is the product of their expectations. (iv) E_1 is moved to the right. (v) μ'_r is here defined as the rth raw moment (moment about the origin) of the true-score distribution.

The quantity following E_2 on the left side of (9) is seen to be an unbiased estimate of $\mu'_1\mu'_2$. An unbiased estimate of any product of true-score moments about the origin may be obtained similarly by Rule 1.

Rule 1. *To get a type-2-unbiased estimate of the product* $\mu'_{r_1}\mu'_{r_2} \cdots \mu'_{r_W}$ of *raw true-score moments,* (i) *replace each* μ'_{r_w} $(w = 1, 2, \cdots, W)$ *by the symbol* π *with* r_w *different subscripts,* (ii) *take the r-tuple sum of the resulting product* of π's *over all* $r = \sum^W r_w$ *subscripts, no two of which may assume the same value,* (iii) *divide by* $n^{[r]}$.

For example, an unbiased estimate for $\mu'_1 = \bar{\zeta}$ in type-2 sampling is

$$\text{Est}_2\, \mu'_1 = \frac{1}{n} \sum_i \pi_i , \tag{10}$$

where Est$_2$ is read "a type-2-unbiased estimate of." Similarly,

$$\text{Est}_2\, \mu'_2 = \frac{1}{n^{[2]}} \sum_{i \neq j}\sum \pi_{ij} , \tag{11}$$

$$\text{Est}_2\, \mu'^2_1 = \frac{1}{n^{[2]}} \sum_{i \neq j}\sum \pi_i\pi_j , \tag{12}$$

$$\text{Est}_2\, \mu'_3\mu'^2_2\mu'^3_1 = \frac{1}{n^{[10]}} \sum_{i_1 \neq i_2 \neq \cdots \neq i_{10}}\sum \cdots \sum \pi_{i_1 i_2 i_3}\pi_{i_4 i_5}\pi_{i_6 i_7}\pi_{i_8}\pi_{i_9}\pi_{i_{10}} . \tag{13}$$

If the number, N, of examinees in the group actually tested is large, the symbol π in Rule 1 and in equations (9) through (13) may be replaced by p to obtain type-2-unbiased estimates from actual sample data.

Matrix-Unbiased Estimates. Rule 1 gives type-2-unbiased estimates of products of true-score moments, expressed in terms of products of π's. If N is not large, it is desirable to find type-1-unbiased estimates of these products of π's. Such estimates are matrix-unbiased estimates of the products of true-score moments, i.e., simultaneously type-1- and type-2-unbiased. In the discussion of such estimates, the assumption that N is large will be dropped. First of all, the two quantities

$$p_i = \frac{1}{N} \sum_{a=1}^{N} x_{ia} , \tag{14}$$

$$p_{ij} = \frac{1}{N} \sum_{a=1}^{N} x_{ia}x_{ja} ,$$

are sample means in type-1 samples and hence are type-1-unbiased estimates of π_i and π_{ij}, respectively:

$$\begin{aligned} \text{Est}_1\, \pi_i &= p_i\ , \\ \text{Est}_1\, \pi_{ij} &= p_{ij}\ . \end{aligned} \tag{15}$$

To take a less simple case, consider the covariance between items i and j,

$$s_{ij} = p_{ij} - p_i p_j\ . \tag{16}$$

According to the usual formula for the expected value of a sample covariance,

$$E_1 s_{ij} = (N - 1)\sigma_{ij}/N, \tag{17}$$

where

$$\sigma_{ij} = \pi_{ij} - \pi_i \pi_j\ . \tag{18}$$

Thus it is seen that

$$\begin{aligned} E_1 \frac{1}{N-1}(Np_i p_j - p_{ij}) &= E_1\left(-\frac{N}{N-1} s_{ij} + p_{ij}\right) \\ &= -\sigma_{ij} + \pi_{ij} = \pi_i \pi_j\ . \end{aligned} \tag{19}$$

Now, since the sample mean of the observed test score is

$$\bar{z} = \frac{1}{n} \sum_i p_i\ , \tag{20}$$

it follows from (10), (15), (14), and (20), that

$$\text{Est}_{12}\, \mu_1' = \text{Est}_1 \frac{1}{n} \sum_i \pi_i = \frac{1}{n} \sum_i p_i = \bar{z}. \tag{21}$$

Similarly, from (11) and (15),

$$\text{Est}_{12}\, \mu_2' = \text{Est}_1 \frac{1}{n^{[2]}} \sum_{i \neq j}\sum \pi_{ij} = \frac{1}{n^{[2]}} \sum_{i \neq j}\sum p_{ij}\ . \tag{22}$$

Likewise, from (12) and (19)

$$\begin{aligned} \text{Est}_{12}\, \mu_1'^2 &= \text{Est}_1 \frac{1}{n^{[2]}} \sum_{i \neq j}\sum \pi_i \pi_j \\ &= \frac{1}{n^{[2]}} \frac{1}{N-1} \sum_{i \neq j}\sum (Np_i p_j - p_{ij}). \end{aligned} \tag{23}$$

Equations (21), (22), and (23) provide estimators of μ_1' , μ_2' , and $\mu_1'^2$ that are unbiased in matrix sampling.

In general, estimates that are type-2-unbiased will differ from estimates that are matrix-unbiased at most by quantities of order $1/N$. Since such quantities will be negligible if N is at all large, further consideration will center on type-2-unbiased estimates.

Computational Formulas for Type-2-Unbiased Estimates. Although the type-2-unbiased estimators so far discussed are exact, these estimators are not yet expressed in a convenient computational form. More convenient formulations will now be obtained, retaining quantities of order $1/n$, since these will be needed for some purposes, but in certain cases neglecting quantities $O(1/n^2)$. The reason for retaining quantities $O(1/n)$ is apparent from the fact that the variance of the observed scores differs from the variance of the true scores only by quantities of this order.

Very large N is again assumed. Further notation will be needed.

rth raw moment of observed proportion-scores:

$$m'_r = \frac{1}{N} \sum_a z_a^r . \tag{24}$$

Variance of item difficulties:

$$s_p^2 = \frac{1}{n} \sum_{i=1}^{n} p_i^2 - \bar{z}^2. \tag{25}$$

rth raw moment of item difficulties:

$$M'_r = \frac{1}{n} \sum_{i=1}^{n} p_i^r . \tag{26}$$

The foregoing quantities are all statistics for the sample of examinees tested and are thus computable from the data.

First of all, z_a has a binomial distribution in type-2 sampling. The factorial moments ([8], pp. 56–60) about the origin of a binomial distribution have a particularly simple form ([8], p. 118): the rth factorial sampling moment of $x_a = nz_a$ is

$$\mu'_{[r]}(x_a) = E_2 x_a^{[r]} = n^{[r]} \zeta_a^r . \tag{27}$$

It is thus clear that

$$\mathrm{Est}_2\, \mu'_r = \frac{1}{n^{[r]}} E_1 x_a^{[r]} \doteq \frac{1}{n^{[r]}} A_a x_a^{[r]}. \tag{28}$$

The quantity on the right is a factorial moment readily computable from the frequency distribution of observed scores. Computing formulas for $r = 1, \cdots, 6$ are given in (35) for convenient reference. Equation (28) may be stated as a theorem.

THEOREM 1. *The rth raw factorial moment of the distribution of number-correct observed scores,* x_a , *divided by* $n^{[r]}$ *is a type-2-unbiased estimate of the rth raw ordinary moment of the distribution of proportion-correct true scores,* ζ_a .

Computational formulas for unbiased estimates of various products of true-score moments will now be obtained. First of all, by Rule 1,

$$(29) \qquad \mathrm{Est}_2\, \mu_1'^r = \frac{1}{n^{[r]}} \sum_{i_1 \neq i_2 \neq} \sum \cdots \sum_{\cdots \neq i_r} \pi_{i_1}\pi_{i_2} \cdots \pi_{i_r} = \frac{r!}{n^{[r]}} (1^r),$$

where (1^r) denotes an rth order unitary symmetric function ([1], p. 431). This symmetric function is readily rewritten in terms of power-sums with the help of the tables of coefficients provided in [2]. The rth power-sum of π_i is n times the rth raw moment of the distribution of the π_i, for which may be substituted the computed value of M_r', the rth raw moment of the distribution of the p_i. The results are given for $r = 2, 3, 4$ in (36). The necessary M's are readily computed for any given set of item data from the frequency distribution of the n observed item difficulties, p_i.

A similar but more complicated procedure may be applied to estimate other products of true-score moments. For example, starting from Rule 1,

$$\begin{aligned}
\mathrm{Est}_2\, \mu_2'\mu_1' &= \frac{1}{n^{[3]}} \sum_{g \neq h \neq i}^{n(n-1)(n-2)} \sum \sum \pi_{gh}\pi_i \\
(30) \qquad &\equiv \frac{1}{n^{[3]}} \Big(\sum^n \sum^n \sum^n \pi_{gh}\pi_i - 2 \sum^n \sum^n \pi_{gi}\pi_i \\
&\qquad - \sum^n \sum^n \pi_{gg}\pi_i + 2 \sum^n \pi_{gg}\pi_g\Big).
\end{aligned}$$

The tables in [2] do not provide the last expression directly, but they do greatly facilitate what would otherwise be an awkward algebraic task. The reader may easily verify the algebraic identity, however, by splitting up the various terms in the last expression, e.g.,

$$\begin{aligned}
\sum^n \sum^n \sum^n \pi_{gh}\pi_i &\equiv \sum_{g \neq h \neq i} \sum \sum \pi_{gh}\pi_i \\
&\qquad + 2 \sum_{g \neq i} \sum \pi_{gi}\pi_i + \sum_{g \neq i} \sum \pi_{gg}\pi_i + \sum \pi_{gg}\pi_g ,
\end{aligned}$$

and then recombining the results. Since $\pi_{gg} = \pi_g$, (30) can be rewritten as shown in the first equation in (31). The process illustrated in (30) becomes an easy one to carry out without the use of tables if all terms $O(1/n^2)$ are dropped. In (30), the term $2 \sum^n \pi_{gg}\pi_g/n^{[3]}$ is $O(1/n^2)$.

Formulas derived similarly are given in (31) for the expected values of every product of true-score moments up to and including those of the fourth order, excepting those covered by (36) and by Theorem 1. The last term of the first equation in (31) after multiplication by $1/n^{[3]}$ is only of order $1/n^2$. All such terms are represented in the remaining equations of (31) by the symbols $O(1/n^2)$.

$$\begin{aligned}
\mathrm{Est}_2\, \mu_1'\mu_2' &= \frac{1}{n^{[3]}} \Big(nM_1' \sum^n \sum^n \pi_{gh} \\
&\qquad - 2 \sum^n \sum^n \pi_{gh}\pi_h - n^2 M_1'^2 + 2nM_2'\Big),
\end{aligned}$$

$$\begin{aligned}
\text{Est}_2\, \mu_1'^2 \mu_2' &= \frac{1}{n^{[4]}} \Big[(n^2 M_1'^2 - n M_2') \sum^n \sum^n \pi_{gh} \\
&\qquad - n^3 M_1'^3 - 4n M_1' \sum^n \sum^n \pi_{gh}\pi_h\Big] + O\Big(\frac{1}{n^2}\Big) , \\
\text{Est}_2\, \mu_2'^2 &= \frac{1}{n^{[4]}} \Big[\Big(\sum^n \sum^n \pi_{gh}\Big)^2 \\
&\qquad - 2n M_1' \sum^n \sum^n \pi_{gh} - 4 \sum^n \sum^n \sum^n \pi_{gh}\pi_{gi}\Big] + O\Big(\frac{1}{n^2}\Big) , \\
\text{Est}_2\, \mu_3' \mu_1' &= \frac{1}{n^{[4]}} \Big(n M_1' \sum^n \sum^n \sum^n \pi_{ghi} \\
&\qquad - 3n M_1' \sum^n \sum^n \pi_{gh} - 3 \sum^n \sum^n \sum^n \pi_{ghi}\pi_i\Big) + O\Big(\frac{1}{n^2}\Big) .
\end{aligned} \tag{31}$$

The computation of the quantities in (31) that involve summation signs can be further simplified, after substituting p's for π's.

$$\sum^n \sum^n p_{gh} = \sum_g \sum_h A_a x_{ga} x_{ha} = A_a\Big(\sum_g x_{ga}\Big)^2 = n^2 m_2' . \tag{32}$$

$$\begin{aligned}
\sum^n \sum^n p_{gh} p_h &= \sum_h p_h \sum_g p_{gh} = \sum_h p_h \sum_g A_a x_{ga} x_{ha} \\
&= n \sum_h p_h A_a x_{ha} z_a = n \sum_h p_h \dot{z}_h ,
\end{aligned} \tag{33}$$

where $\dot{z}_h = A_a x_{ha} z_a$ is $1/N$ times the sum of the scores of those examinees who answered item h correctly. Similarly

$$\begin{aligned}
\sum^n \sum^n \sum^n p_{gh} p_{gi} &= n^2 \sum_g \dot{z}_g^2 , \\
\sum^n \sum^n \sum^n p_{ghi} &= n^3 m_3' , \\
\sum^n \sum^n \sum^n p_{ghi} p_i &= n^2 \sum p_i \ddot{z}_i ,
\end{aligned} \tag{34}$$

where $\ddot{z}_i = A_a x_{ia} z_a^2$ is $1/N$ times the sum of the squared scores of those examinees who answered item i correctly.

Equations (35), (36), and (37) summarize for convenient reference all computational formulas for type-2-unbiased estimates of products of true-score raw moments up through the fourth order. Equations (35), given through the sixth order, are obtained from (28) and from the standard formulas ([8], eq. 3.18) expressing factorial moments in terms of the ordinary raw moments. Equations (36) are obtained as described in connection with (29). Equations (37) are obtained by substituting (32), (33), (34) in (31) and dropping terms of order $1/n^2$.

$$
\begin{aligned}
\mathrm{Est}_2\, \mu_1' &= m_1' , \\
\mathrm{Est}_2\, \mu_2' &= \frac{1}{n^{[2]}} (n^2 m_2' - n m_1'), \\
\mathrm{Est}_2\, \mu_3' &= \frac{1}{n^{[3]}} (n^3 m_3' - 3n^2 m_2' + 2n m_1'), \\
\mathrm{Est}_2\, \mu_4' &= \frac{1}{n^{[4]}} (n^4 m_4' - 6n^3 m_3' + 11n^2 m_2' - 6n m_1'), \\
\mathrm{Est}_2\, \mu_5' &= \frac{1}{n^{[5]}} (n^5 m_5' - 10n^4 m_4' \\
&\qquad + 35n^3 m_3' - 50n^2 m_2' + 24n m_1'), \\
\mathrm{Est}_2\, \mu_6' &= \frac{1}{n^{[6]}} (n^6 m_6' - 15n^5 m_5' \\
&\qquad + 85n^4 m_4' - 225n^3 m_3' + 274n^2 m_2' - 120n m_1').
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
\mathrm{Est}_2\, \mu_1'^2 &= \frac{1}{n^{[2]}} (n^2 M_1'^2 - n M_2'), \\
\mathrm{Est}_2\, \mu_1'^3 &= \frac{1}{n^{[3]}} (n^3 M_1'^3 - 3n^2 M_1' M_2' + 2n M_3'), \\
\mathrm{Est}_2\, \mu_1'^4 &= \frac{1}{n^{[4]}} (n^4 M_1'^4 - 6n^3 M_2' M_1'^2 + 3n^2 M_2'^2 + 8n^2 M_3' M_1' - 6n M_4').
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
\mathrm{Est}_2\, \mu_1' \mu_2' &\doteq \frac{1}{n^{[3]}} (n^3 m_1' m_2' - 2n \sum_i p_i \dot{z}_i - n^2 m_1'^2), \\
\mathrm{Est}_2\, \mu_1'^2 \mu_2' &\doteq \frac{1}{n^{[4]}} [n^3 (n m_1'^2 - M_2') m_2' - n^3 m_1'^3 - 4n^2 m_1' \sum_i p_i \dot{z}_i], \\
\mathrm{Est}_2\, \mu_2'^2 &\doteq \frac{1}{n^{[4]}} (n^4 m_2'^2 - 2n^3 m_1' m_2' - 4n^2 \sum_i \dot{z}_i^2), \\
\mathrm{Est}_2\, \mu_3' \mu_1' &\doteq \frac{1}{n^{[4]}} (n^4 m_1' m_3' - 3n^3 m_1' m_2' - 3n^2 \sum_i p_i \ddot{z}_i).
\end{aligned}
\tag{37}
$$

[*Computational Note.* If computations will involve (37) as well as (35) and (36), only the first two terms of each polynomial on the right should be used from (35) and (36). Equations (37) do not retain terms $O(1/n^2)$, and it seems to be better here to discard all terms $O(1/n^2)$ rather than to retain some when others are discarded.]

Estimating the True-Score Central Moments. For many purposes, type-2-unbiased estimates of the true-score central (about the mean) moments, rather than of the true-score raw (about the origin) moments, are desired. These are obtained by expressing the central moments in terms of products of the raw moments ([8], p. 50) and then replacing these products by their unbiased estimators. Thus a type-2-unbiased estimate of the true-score variance is

$$\begin{aligned} \text{Est}_2\, \sigma_\zeta^2 &= \text{Est}_2\, (\mu_2' - \mu_1'^2) = \frac{1}{n^{[2]}} [A_a x_a(x_a - 1) - n^2 M_1'^2 + n M_2'] \\ &= \frac{1}{n-1} [n s_z^2 - \bar{z}(1 - \bar{z}) + s_p^2]. \end{aligned} \tag{38}$$

A matrix-unbiased estimate of the true-score variance can be obtained from (22), (23), and (16) if desired:

$$\begin{aligned} \text{Est}_{12}\, \sigma_\zeta^2 &= \frac{1}{n^{[2]}} \frac{N}{N-1} \sum_{i \neq j}\sum (p_{ij} - p_i p_j) \\ &= \frac{N}{N-1} \frac{1}{n^{[2]}} \left(\sum^n \sum^n s_{ij} - \sum^n p_i + \sum^n p_i^2\right) \\ &= \frac{N}{N-1} \text{Est}_2\, \sigma_\zeta^2 \,. \end{aligned} \tag{39}$$

It is interesting to note that if test reliability is defined by the equation

$$r_{zz} = r_{xx} = \frac{\text{Est}_2\, \sigma_\zeta^2}{\sigma_z^2}, \tag{40}$$

then substitution from (38) shows that the test reliability is the same as the Kuder-Richardson formula-20 reliability coefficient, r_{20}, as is conveniently seen from ([14], eq. 27).

Estimating the Shape of the True-Score Distribution

An estimated frequency distribution of true scores may be obtained in a variety of ways from the moments. If only the first four true-score moments have been estimated, Pearson-type curves may be fitted by the formulas given in [3]. If more than four moments are available, a Gram-Charlier or an Edgeworth series may be fitted ([8], pp. 147–156).

The frequency distribution may be fitted to the estimated raw moments that are readily obtained from (35). Better results will presumably be obtained, however, by fitting the distribution to the estimated central moments obtained by the method of the immediately preceding section.

Estimating the True Score of Each Examinee

Suppose that the psychologist is continually testing many groups of examinees with randomly parallel forms of some test. Suppose that he makes an estimate, $\hat{\zeta}_a$, of the true score of each examinee. Over a period of time he wishes to minimize the sum of squares of the discrepancies between his estimates and the true values—he wishes to minimize $E_{12}(\hat{\zeta}_a - \zeta_a)^2$. (The problem here is basically similar to that discussed by Robbins in ([12], p. 161), but there are important differences and the method of estimation used here is different from his.)

Linear Estimation of True Score from Observed Score

In the simplest case, it is assumed that the estimate is a linear function of observed score. In this case, the problem is to minimize

$$S_1 = E_{12}(Bz_a + C - \zeta_a)^2$$

by a suitable choice of B and C, which are assumed to be functions of the true scores, not of the observed scores.

Since in type-2 sampling z_a is actually a sufficient estimator of ζ_a, it might be thought that S_1 would be minimized by setting $B = 1$ and $C = 0$. This is not the case, however.

The necessary derivatives are

$$\frac{\partial S_1}{\partial B} = 2E_{12}z_a(Bz_a + C - \zeta_a), \tag{41}$$

$$\frac{\partial S_1}{\partial C} = 2E_{12}(Bz_a + C - \zeta_a). \tag{42}$$

Set (42) equal to zero, take expected values, and solve for C to obtain

$$C = \bar{\zeta}(1 - B). \tag{43}$$

Substitute (43) in (41) to obtain

$$(E_{12}z_a^2 - \bar{\zeta}E_{12}z_a)B = E_{12}z_a\zeta_a - \bar{\zeta}E_{12}z_a\,. \tag{44}$$

By (28),

$$E_{12}z_a^2 = \frac{1}{n^2}E_{12}(x_a^{[2]} + x_a) = \frac{n-1}{n}\mu_2' + \frac{1}{n}\mu_1'\,. \tag{45}$$

Since $E_{12}z_a\zeta_a = E_1\zeta_aE_2z_a = \mu_2'$, (44) becomes

$$
\begin{aligned}
B &= \frac{\mu_2' - \mu_1'^2}{\dfrac{n-1}{n}\,\mu_2' + \dfrac{1}{n}\,\mu_1' - \mu_1'^2} = \frac{n\sigma_\zeta^2}{(n-1)\sigma_\zeta^2 + \bar{\zeta}(1-\bar{\zeta})} \\
&= \frac{n}{n-1}\left[1 - \frac{\bar{\zeta}(1-\bar{\zeta})}{(n-1)\sigma_\zeta^2 + \bar{\zeta}(1-\bar{\zeta})}\right].
\end{aligned}
\tag{46}
$$

Since $\bar{\zeta}$, $\bar{\zeta}^2$, and σ_ζ^2 are unknown, they must be replaced by sample estimates. These will have a type-2 sampling variance of order $1/n$, and will therefore probably differ from the true values by an amount of order $1/\sqrt{n}$. Such a substitution into the last expression in (46) will change the value of B only by $O(n^{-3/2})$, because of the $n - 1$ in the denominator of the second fraction. If unbiased estimates from (35), (36), and (38) are used, the estimated value of B is found to be

$$
\hat{B} = \frac{n}{n-1}\left[1 - \frac{\bar{z}(1-\bar{z}) + s_p^2/(n-1)}{ns_z^2 + ns_p^2/(n-1)}\right].
\tag{47}
$$

This quantity differs from the Kuder-Richardson formula-21 reliability coefficient ([4], p. 225)

$$
r_{21} = \frac{n}{n-1}\left[1 - \frac{\bar{z}(1-\bar{z})}{ns_z^2}\right]
\tag{48}
$$

only by quantities of order $1/n^2$:

$$
\begin{aligned}
r_{21} &- \frac{n}{n-1}\left[1 - \frac{\bar{z}(1-\bar{z}) + s_p^2/(n-1)}{ns_z^2 + ns_p^2/(n-1)}\right] \\
&= \frac{s_p^2[s_z^2 - \bar{z}(1-\bar{z})]}{(n-1)s_z^2[(n-1)s_z^2 + s_p^2]} = O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Consequently, r_{21} may be used as an estimator for B. It should be noted that r_{21} and r_{20} differ by $O(1/n)$. Hence r_{20} is *not* an adequate estimator of B in the present situation, since r_{21} also differs from 1.00 by $O(1/n)$.

The linear equation for estimating an examinee's true score from his observed score is thus seen to be

$$
\hat{\zeta}_a = \bar{z} + \hat{B}(z_a - \bar{z}) \quad \text{or} \quad \hat{\xi}_a = \bar{x} + \hat{B}(x_a - \bar{x}).
\tag{49}
$$

This last equation is the same as (1) except that $\hat{B}$ now appears in place of r_{xx}.

Curvilinear Estimation of True Score from Observed Score

A virtue of the present approach is that it can be extended to the case where the regression of true score on observed score is curvilinear. This case is likely to arise when the true-score distribution in the group tested is substantially different from the usual symmetrical bell-shaped form. This situation will arise not only whenever the group of examinees is a peculiar one, but also whenever the test administered is too hard or too easy for the group.

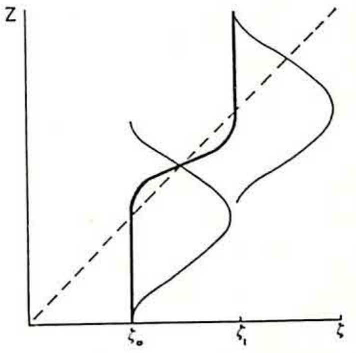


FIGURE 1

Regression (heavy line) of true score (ζ) on obtained score (z) when true score has a dichotomous frequency distribution

Figure 1 presents an extreme example to illustrate why a rectilinear relation cannot always be expected. The situation in Figure 1 is that half of the group have true scores at ζ_0 and half at ζ_1 . The conditional frequency distribution of the observed scores for each of the two true-score values is indicated by a bell-shaped curve. Since the average error of measurement is zero, the mean of each conditional frequency distribution lies on the 45-degree broken line, which is the regression of observed score on true score. It is graphically evident from the figure that the other regression, the regression of true score on observed score, must lie very close to the heavy ogive-shaped line and must be very sharply curvilinear.

If the curvilinear regression of true score on observed score can be approximated by a rth degree polynomial in z_a , the equation of this curvilinear regression can be obtained by finding the values of the B's that will minimize $S_r = E_{12}(B_0 + B_1z_a + B_2z_a^2 + \cdots + B_rz_a^r - \zeta_a)^2$ and by writing the regression equation as

$$\hat{\zeta}_a = \hat{B}_0 + \hat{B}_1z_a + \hat{B}_2z_a^2 + \cdots + \hat{B}_rz_a^r , \tag{50}$$

where the $\hat{B}$'s are sample estimates of the B's.

When the partial derivative of S_r with respect to each B is set equal to zero, there results a set of r linear equations in r unknowns which are basically the same as the "normal" equations used to determine curvilinear regression coefficients in conventional situations ([8], eq. 22.20; [11], pp. 429–431). For example, when $r = 3$, the equations found by differentiating S_3 are

$$\begin{aligned} E_{12}\zeta_a &= B_0 + B_1E_{12}z_a + B_2E_{12}z_a^2 + B_3E_{12}z_a^3 , \\ E_{12}\zeta_az_a &= B_0E_{12}z_a + B_1E_{12}z_a^2 + B_2E_{12}z_a^3 + B_3E_{12}z_a^4 , \\ E_{12}\zeta_az_a^2 &= B_0E_{12}z_a^2 + B_1E_{12}z_a^3 + B_2E_{12}z_a^4 + B_3E_{12}z_a^5 , \\ E_{12}\zeta_az_a^3 &= B_0E_{12}z_a^3 + B_1E_{12}z_a^4 = B_2E_{12}z_a^5 + B_3E_{12}z_a^6 . \end{aligned} \tag{51}$$

In order to evaluate $E_{12}z_a^r$, it is first convenient to express z_a^r as a sum of the polynomials $x_a^{[r']}$ ($r' = 1, 2, \cdots, r$), as in (45). The expected values of each of these polynomials is then readily obtained from (28). Thus,

$$
\begin{aligned}
E_{12}z_a^2 &= \frac{1}{n^2}(n^{[2]}\mu_2' + n\mu_1'),\\
E_{12}z_a^3 &= \frac{1}{n^3}E_{12}(x_a^{[3]} + 3x_a^{[2]} + x_a^{[1]}),\\
&= \frac{1}{n^3}(n^{[3]}\mu_3' + 3n^{[2]}\mu_2' + n\mu_1'),\\
E_{12}z_a^4 &= \frac{1}{n^4}(n^{[4]}\mu_4' + 6n^{[3]}\mu_3' + 7n^{[2]}\mu_2' + n\mu_1'), \qquad (52)\\
E_{12}z_a^5 &= \frac{1}{n^5}(n^{[5]}\mu_5' + 10n^{[4]}\mu_4' + 25n^{[3]}\mu_3' + 15n^{[2]}\mu_2' + n\mu_1'),\\
E_{12}z_a^6 &= \frac{1}{n^6}(n^{[6]}\mu_6' + 15n^{[5]}\mu_5' + 65n^{[4]}\mu_4'\\
&\qquad + 90n^{[3]}\mu_3' + 31n^{[2]}\mu_2' + n\mu_1').
\end{aligned}
$$

The numerical coefficients on the right are the same as those obtained when an ordinary raw moment is expressed as a sum of factorial moments ([8], eq. 3.19).

Similarly,

$$
\begin{aligned}
E_{12}\zeta_a z_a &= \mu_2',\\
E_{12}\zeta_a z_a^2 &= \frac{1}{n^2}(n^{[2]}\mu_3' + n\mu_2'), \qquad (53)\\
E_{12}\zeta_a z_a^3 &= \frac{1}{n^3}(n^{[3]}\mu_4' + 3n^{[2]}\mu_3' + n\mu_2').
\end{aligned}
$$

It is helpful to rewrite (51) by making the substitutions $C_0 = nB_0$, $C_1 = n(B_1 - 1)$, $C_2 = nB_2$, $C_3 = nB_3$. With a little rearranging, (51) becomes

$$
\begin{aligned}
C_3E_{12}z_a^3 + C_2E_{12}z_a^2 + C_1E_{12}z_a + C_0 &= n(E_{12}\zeta_a - E_{12}z_a) = 0,\\
C_3E_{12}z_a^4 + C_2E_{12}z_a^3 + C_1E_{12}z_a^2 + C_0E_{12}z_a &= n(E_{12}\zeta_a z_a - E_{12}z_a^2), \qquad (54)\\
C_3E_{12}z_a^5 + C_2E_{12}z_a^4 + C_1E_{12}z_a^3 + C_0E_{12}z_a^2 &= n(E_{12}\zeta_a z_a^2 - E_{12}z_a^3),\\
C_3E_{12}z_a^6 + C_2E_{12}z_a^5 + C_1E_{12}z_a^4 + C_0E_{12}z_a^3 &= n(E_{12}\zeta_a z_a^3 - E_{12}z_a^4).
\end{aligned}
$$

The reason for preferring (54) to (51) is that three of the four unknowns (B's) in (51) are quantities $O(n^{-1})$ and the fourth differs from unity by a quantity $O(n^{-1})$; in (54), all the unknowns are $O(1)$, i.e., are functionally independent of the value of n except for terms that become small as n becomes large. Furthermore, when (52) and (53) are substituted into the right side of (54) all quantities $O(n)$ cancel leaving only quantities $O(1)$.

Adequate approximations to the solutions of (54) may now be obtained by substituting unbiased estimates for the unknown values of the μ' in (52) and (53), substituting the resulting numerical quantities back into (54), and solving the resulting simultaneous linear equations for C_0, C_1, C_2, and C_3. The exact extent to which terms of higher order in n can be neglected in this process is hard to state for any given value of n. There are good indications, however, that terms of order n^{-1} should be retained in (52) and (53) even when n is as large as 100.

REFERENCES

[1] David, F. N. and Kendall, M. G. Tables of symmetric functions, Part I. *Biometrika*, 1949, **36**, 431-449.

[2] David, F. N. and Kendall, M. G. Tables of symmetric functions, Part V. *Biometrika*, 1955, **42**, 223-242.

[3] Elderton, W. P. *Frequency curves and correlation.* (3rd ed.) Cambridge: Cambridge Univ. Press, 1938.

[4] Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.

[5] Hooke, R. Symmetric functions of a two-way array. *Ann. math. Statist.*, 1956, **27**, 55-79.

[6] Hooke, R. Some applications of bipolykays to the estimation of variance components and their moments. *Ann. math. Statist.*, 1956, **27**, 80-98.

[7] Kelley, T. L. *Fundamentals of statistics.* Cambridge: Harvard Univ. Press, 1947.

[8] Kendall, M. G. *The advanced theory of statistics.* (5th ed.) New York: Hafner, 1952.

[9] Lord, F. M. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 1955, **20**, 1-22.

[10] Lord, F. M. Inferring the shape of the frequency distribution of true scores and of errors of measurement. Princeton: Educ. Test. Serv. Res. Bull. 57-9, 1957.

[11] Peters, C. C. and Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.

[12] Robbins, H. An empirical Bayes approach to statistics. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: Univ. California Press, 1956. Vol. 1, pp. 157-163.

[13] Robson, D. S. Applications of multivariate polykays to the theory of unbiased ratio-type estimators. *J. Amer. statist. Ass.*, 1957, **52**, 511-522.

[14] Tucker, L. R. A note on the estimation of test reliability by the Kuder-Richardson Formula (20). *Psychometrika*, 1949, **14**, 117-119.

[15] Tukey, J. W. Some sampling simplified. *J. Amer. statist. Ass.*, 1950, **45**, 501-519.

[16] Tukey, J. W. Keeping moment-like sampling computations simple. *Ann. math. Statist.*, 1956, **27**, 37-54.

Manuscript received 3/3/58

Revised manuscript received 6/13/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

A STATISTICAL TEST FOR THE SIGNIFICANCE OF A COEFFICIENT OF REPRODUCIBILITY

PHILIP C. SAGI

PRINCETON UNIVERSITY

Under the assumption that responses to different dichotomous items are statistically independent, exact distributions for coefficients of reproducibility, CR, are derived. These distributions are useful in testing whether an observed coefficient of reproducibility differs significantly from chance expectations and whether further scaling manipulations are warranted. The effects of purification on sampling distributions of errors are calculated to demonstrate a relationship between scaling operations and expected chance values of the CR.

The practice of devising a measuring instrument on the very sample that is being used to test a substantive hypothesis involving the measure is suspect. The permissive nature of scalogram analysis, the collapsing of item categories, the elimination of items if they contribute too much error, and the ad hoc determination of positive and negative response categories capitalize on the unique and chance features of the sample. The more such operations are performed in the process of scaling the more suspect is the resultant claim of a scale. At some point evidence should be presented which demonstrates (i) that scaling operations were warranted, or (ii) that the parallelogram formed by the response patterns to the retained set of ordered items is not merely an artifact of the operations performed in the analysis and the peculiarities of the sample—but that the items do in fact form a scale with the claimed theoretical properties. Such evidence may be gained in either of two ways—by replication on a subsequent sample or by the use of appropriate tests of statistical significance. Test by replication is neither convenient nor possible at all times.

One weakness of current scaling practices in sociological and psychological research is due to the lack of knowledge about sampling distributions of evaluative summary statistics, such as Guttman's Coefficient of Reproducibility, CR. The intent here is to present exact distributions and exact tests for the CR under certain conditions and to discuss tests of significance for other evaluative statistics. This is not an original problem. Guttman provides an estimate of the upper bound to the standard error of the CR, given certain assumptions ([3], p. 77). This estimate is based upon an intuitive appeal to the "sampling distribution of ordinary proportions (which follow the binomial distribution) especially when the population reproducibility is high (say, over .90)" ([4], p. 279). Loevinger recognizes the existence of the problem

([6], p. 37). The current practice of reporting the chance value of the CR, under the assumption of complete independence, implicitly assumes the existence of a sampling distribution of the CR. Further evidence of the type mentioned can be readily cited.

Reproducibility, Homogeneity, and the Concept of Error

Loevinger and Guttman have approached the problem of measurement in a manner that emphasizes the patterns of responses to a given set of items. In the case of dichotomous items with common content, as judged by face validity, items are ordered by their difficulty. The ranking by difficulty is equivalent to the ranking of items by frequencies of positive endorsements. The pattern of responses to ordered items is then compared to an ideal pattern of responses. Loevinger's Coefficient of Homogeneity and Guttman's CR are summary indices designed to reflect the totality of divergences from ideal patterns. If a response pattern for items ordered from "hard" to "easy" is + − + + − − +, Loevinger would assign the number 7 as a measure of divergence and Guttman would assign the number 3. The numbers 7 and 3 describe the number of errors that would be associated with the observed pattern by the respective systems of measuring divergences. It is clear that the definition of error is somewhat arbitrary and is imposed on the data. Working rules for counting errors, when dichotomous items are ordered from hard to easy, are, respectively:

I. Loevinger—the number of (+ −) patterns among all pairs of items;

II. Guttman—the minimum number of substitutions of + for − or − for + in order to transform the observed pattern into some ideal pattern or scale type.

The two rules have a great deal in common. Each involves the counting of (+ −) patterns. The rule for Guttman scales may be viewed as the counting of (+ −) patterns for adjacent items plus double and triple to n-order error patterns among adjacent items. The (+ + − −) pattern is a double error of the (+ −) pattern, and the (+ + + − − −) pattern is a triple error of this (+ −) pattern. The illustrative pattern + − + + − − + is recognized as composed of a single and a double error pattern, which total to three errors.

Additional arbitrary rules for counting errors can be readily introduced. As an example, and because it appears fruitful, consider a rule III and count as errors the (+ −) patterns among *adjacent* pairs of ordered items. (Green [2] anticipated rule III with minor variations. See [8] for additional rules.) Thus, + − + + − − + would be assigned the number 2 as a measure of divergence. The observation that this arbitrary rule for counting errors gives the same results as does the Guttman system of counting errors when the number of dichotomous items is three or less and agrees closely when the number of dichotomous items is four, suggests the utility of some form of four-fold analysis involving each pair of adjacent items.

Exact Distribution of Errors under Rule III

The expression $p' = (r_1!r_2!c_1!c_2!)/(a!b!c!d!n!)$ is well known and is the probability of the observed four-fold distribution under the hypothesis of statistical independence. Row and column totals are treated as fixed and not subject to sampling variation. This is also a problem in sampling without replacement ([5], pp. 230-231).

Given K items ordered by difficulty from hard to easy, the probability of the observed response patterns, assuming statistical independence, for each adjacent pair of items is

$$\prod_{i=1}^{K-1} p'_i \,.$$

This follows directly from the assumption of independence, where i denotes the four-fold table formed by the ith and $i + 1$ item ordered by difficulty from $i = 1, 2, \cdots, K$. Table 1 is an example of the typical four-fold table formed by adjacent items in which c_i is the frequency of $(+\ -)$ patterns and is therefore the frequency of errors. (In the case of a perfect scale, the c_i cell contains a zero frequency and therefore $\sum_{i=1}^{K-1} c_i = 0$.) In fact

TABLE 1

Typical Four-fold Formed by Adjacent Items

Item i + 1 response categories	Item i response categories		
	+	-	
+	a_i	b_i	$r_{1(i+1)}$
-	c_i	d_i	$r_{2(i+1)}$
	c_{1i}	c_{2i}	n

$$\prod_{i=1}^{K-1} p'_i$$

is the probability of exactly $c_1, c_2, \cdots, c_{K-1}$ errors occurring under the hypothesis of statistical independence. Referring to Table 1 while employing Fisher's exact test, the probability of obtaining a frequency in the c_i cell less than or equal to c_i is given by the expression $p_i = p'_{0i} + p'_{1i} + p'_{2i} + \cdots + p'_{c_i i}$, and the probability of the total errors being less than or equal to E can be expressed as

8-377
7.9.70

$$\text{Prob}\,(c_1 + c_2 + c_3 + \cdots + c_{K-1} \le E) \le \sum_{c_i} \left(\prod_{i=1}^{K-1} p'_{c_i i} \right), \tag{1}$$

where $\sum_{c_i}$ symbolizes the summation for all possible c_i such that

$$\sum_{i=1}^{K-1} c_i \le E.$$

Exact Distribution of Guttman's Coefficient of Reproducibility

Restricting the argument to four dichotomous items, the probability of exactly c_1, c_2, c_3, G, observed errors among four ordered items when errors are counted in keeping with the rule covering Guttman scales is

$$\frac{\prod_{i=1}^{3} p'_{c_i i} \sum_{L=0}^{b} \binom{m_2 - m_1 + c_1}{c_2 - L} \binom{m_1 - c_1}{L} \binom{L}{G} \binom{n_3 - L}{n_4 - c_3 - G}}{\binom{m_2}{c_2} \binom{n_3}{n_4 - c_3}}, \tag{2}$$

where b is equal to the smallest of the three frequencies n_3, c_2, or $m_1 - c_1$ (since $G \le n_3$, $G \le c_2$ and $G \le m_1 - c_1$); m_i is the frequency of positive responses to the ith item; n_i is the frequency of negative responses to the ith item; G is the number of double error patterns. The product factor in (2) is identical to the product factor in (1). The remaining factors of (2) may be viewed as a correction for double errors. Formula (2) is readily apparent if for each particular array of errors c_1, c_2, c_3, there exists a sampling distribution of double error patterns distributed in the same manner as (+ −) patterns for three adjacent items.

The Guttman rule for counting errors introduces an impressive complexity to the formula previously derived and a still greater complexity for Loevinger's rule. While it is possible to develop similar formulas for K items, it is not practical to do so. The practical solution may be to use rule III for the measurement of divergences.

In Tables 2 and 3 exact distributions are presented for errors counted according to Guttman and rule III. Table 2 compares exact distributions for the two systems of counting errors given a sample of size 10, four items, and m_1, m_2, m_3, m_4, equal to 2, 4, 6, 8, respectively. Table 3 compares exact distributions for the two systems of counting errors given a sample size $N = 10$, four items ($K = 4$), and all item marginals equal to five ($m_i = 5$). Where marginals are all equal, items cannot be ordered by marginals. However, we can assume either an a priori ordering or marginals that differ by extremely small numbers. These parameter values were chosen for illustrative purposes and ease of computation.

The columns headed cumulative percentages provide the information for one-tailed tests of significance for the evaluative statistic CR. Specifically, the cumulative percentages indicate the chances of getting a CR as large or

TABLE 2

Exact Distribution of E
($N=10$, $m_1=2$, $m_2=4$, $m_3=6$, $m_4=8$)

Number of errors E	Guttman's system		Rule III		CR
	Hypothetical frequencies	Cum. %	Hypothetical frequencies	Cum. %	
0	2,520	00.13	2,520	00.13	1
1	30,240	01.65	33,600	01.82	.975
2	158,340	09.63	175,560	10.67	.95
3	431,424	31.37	457,632	33.73	.925
4	638,470	63.54	634,494	65.70	.900
5	502,432	88.86	472,416	89.50	.875
6	190,792	98.47	178,668	98.51	.850
7	29,232	99.95	28,560	99.95	.825
8	1,050	100.00	1,050	100.00	.80
Total	1,984,500		1,984,500		

TABLE 3

Exact Distribution of E
($N = 10$, $m_i = 5$; $i = 1, 2, 3, 4$)

Number of errors E	Guttman's system		Rule III		CR
	Hypothetical frequencies	Cum. %	Hypothetical frequencies	Cum. %	
0	1	.0000	1	.0000	1
1	50	.0003	75	.0005	.975
2	1,100	.0072	2,175	.0141	.95
3	13,825	.0936	30,925	.2073	.925
4	98,150	.7069	232,575	1.6606	.900
5	417,252	3.3142	1,001,253	7.9173	.875
6	1,544,200	12.9636	2,592,025	24.1144	.850
7	3,012,150	31.7860	4,142,475	50.0000	.825
8	4,516,175	60.0068	4,142,475	75.8856	.80
9	3,726,350	83.2921	2,592,025	92.0827	.775
10	2,028,879	95.9702	1,001,253	98.3394	.750
11	549,200	99.4021	232,575	99.7927	.725
12	88,925	99.9578	30,925	99.9859	.70
13	6,550	99.9987	2,175	99.9995	.675
14	200	100.0000	75	100.0000	.650
15	1	100.0000	1	100.0000	.625
Total	16,003,008		16,003,008		

larger than the one observed if the hypothesis that responses to items are statistically independent is true. Rejection of the null hypothesis in favor of the alternative hypothesis, the hypothesis that the CR is greater than values usually observed if the null hypothesis is true, is taken as evidence that the item response patterns are in the direction of a scale. Thus, assuming statistical independence and given the parameter values as indicated in Table 2, values of CR greater than or equal to .90 occur with probability less than or equal to .6570 or .6354, depending on the system of counting errors. By definition

$$\mathrm{CR} = 1 - \frac{E}{NK}.$$

In Table 3, the probabilities corresponding to a CR of .90 are .016606 and .007069. The differences between the probabilities indicate the similarity of results between the two systems of counting errors (at least for four-item scales) and the important effect of the item marginals on the distributions. In Table 4 the values of E counted in accord with rule III for given probability levels under the hypothesis of statistical independence are tabulated. The values of N, K, and m_i respectively, are 100, 4, and 20, 40, 60, 80. Comparisons between Tables 2 and 4 provide the additional and expected information that the size of the sample is also an important consideration.

While the shape of the sampling distribution varies with N, K, and m_i ($i = 1, 2, \cdots, K$), the distributions approximate normality when N becomes large. Goodman [1] has demonstrated this latter point and has shown the adequacy of the approximation for a small sample problem. For small samples, the approximation seems most warranted when the m_i/N are all about .5. In the small sample case, radical departures from approximate normality obtain with extreme item marginals. Exact distributions, if known over a range of values for N, K, and m_i, may delimit the situations in which Goodman's approximate method yields accurate probability statements. The reader is referred to [1] for a more adequate treatment of the large sample problem.

Discussion

Under the conditions described, exact tests are generally applicable if errors are counted according to rule III. For the Guttman rule, we are restricted to 2, 3, and 4-item scales. There are practical limitations that must be considered. When N is large, greater than 10, and K is 4 or greater, the computational labor involved in the use of the derived formulas becomes excessive. The utility of the formulas is further limited by the logical argument that the correct application of the tests occur prior to "purification," the practice of eliminating items which contribute the greatest numbers of errors. The common practice in scaling is to purify and then report the

TABLE 4

Exact Distribution of Errors
($N=100$, $m_1=20$, $m_2=40$, $m_3=60$, $m_4=80$)

Number of errors E	Relative frequencies	Cumulative percentages	CR
24	.0000006	.000+	.9400
25	.0000277	.003	.9375
26	.0000777	.011	.9350
27	.0002203	.033	.9325
28	.0005522	.088	.9300
29	.0012905	.217	.9275
30	.0027770	.495	.9250
31	.0055665	1.051	.9225
32	.0103488	2.086	.9200
33	.0179269	3.879	.9175
34	.0288232	6.761	.9150

evaluative statistic calculated for the purified scale. Obviously, a test of the purified-scale statistic must take into consideration the number and marginals of all items *prior to* elimination. Formulas (1) and (2) may be used to test whether a set of items is worth purifying and scaling. The importance of the argument cannot be understated. Even if the hypothesis of statistical independence is true for responses to some original K' items, it is likely that purification will yield a subset of K items ($K < K'$) that appear to scale.

TABLE 5

Exact Distributions of E Following Purification
($N = 10$, $K = 4$, $m_i = 5$)

Number of errors E	Guttman's system		Rule III		CR
	Hypothetical frequencies	Cum. %	Hypothetical frequencies	Cum. %	
0	1	.000	1	.000	1.000
1	50	.003	75	.004	.975
2	1,100	.058	2,175	.113	.950
3	13,625	.739	30,625	1.643	.900
4	90,000	5.243	217,500	12.516	.875
5	304,000	20.440	750,000	50.009	.875
6	871,500	64.007	1,000,000	100.000	.850
7	720,000	100.000			.825
Total	2,000,376		2,000,376		

Had formulas (1) and (2) been applied prior to purification, the tests of significance would have indicated, with predetermined probabilities of error, that purification was or was not justified. Illustrative of this point is the effect that purification has upon the expected chance value of the CR as well as the sampling distribution of E.

Assuming statistically independent responses to different ordered dichotomous items, the expected frequencies, c_i , for any pair of adjacent items may be computed by the formula $m_i n_{i+1}/N$. If, among other equally arbitrary standards, the criterion of purification employed is the elimination of an item or items yielding observed frequencies $c_i \geq m_i n_{i+1}/N$, the remaining items must exhibit frequencies $c_i < m_i n_{i+1}/N$. Further, if it is assumed that the original pool of items is sufficiently large, it is inevitable that some subset of K items will satisfy the purification criterion. Exact distributions for such purified scales are presented in Table 5. Distributions were calculated from (1) and (2) subject to the restriction, $c_i < m_i n_{i+1}/N$. Purification tends to increase the magnitude of the expected CR under the hypothesis of statistical independence.

The expected CR's following purification are .866 and .848 as compared to the unpurified scale statistics of .812 and .797, respectively. Expected chance values of CR that are commonly reported do not correspond to the operations performed on data and actually underestimate the true expected chance value of the CR. Moreover, applications of formulas (1) and (2) to purified sets of items increase the chances of falsely rejecting null hypotheses. Had some other criterion of purification been invoked, a different distribution of errors would have followed.

Other limitations on the use of these formulas also involve operations commonly associated with the scaling process. The stipulation that positive and negative response categories be determined by face validity prior to the inspection of inter-item relationships is not unlike the restriction that formulas (1) and (2) are appropriately used prior to purification. The ad hoc procedure of determining sign by examining the direction of association with other items under consideration also tends to result in observed c_i being less than or equal to expected c_i frequencies. Similarly, the practice of collapsing item-response categories to form contrived dichotomous items is not an admissible operation if such collapsing is done on the basis of observed inter-item response patterns.*

The conditions under which tests of significance apply define prior admissible scaling operations. Conversely, different scaling operations usually imply different sampling distributions of the evaluative statistic. The implication is immediately apparent: either scaling operations are restricted to those operations consistent with developed statistical tests, or extensive replication is used as a substitute testing procedure. Tests proposed in this

*Interestingly enough, such ad hoc collapsing exploits any lack of an isotropic relationship between adjacent items, a situation that is inimical to the concept of a scale.

paper and in [1] merely serve to establish whether responses to a set of ordered dichotomous items are worthy of further manipulation. The complete rationalization of the scaling process requires a test designed to differentiate between artifact and true improvement of the CR following scaling manipulations.

REFERENCES

[1] Goodman, L. A. Simple statistical methods for scalogram analysis. *Psychometrika,* 1959, **24,** 29-43.

[2] Green, B. F. A method of scalogram analysis using summary statistics. *Psychometrika,* 1956, **21,** 79-88.

[3] Guttman, L. The basis for scalogram analysis. In S. A. Stouffer et al., *Measurement and prediction.* Princeton: Princeton Univ. Press, 1950. Pp. 60-90.

[4] Guttman, L. Problems of reliability. In S. A. Stouffer et al., *Measurement and prediction.* Princeton: Princeton Univ. Press, 1950. Pp. 277-311.

[5] Kenney, J. F. and Keeping, E. S. *Mathematics of statistics,* Pt. II. New York: Van Nostrand, 1951.

[6] Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.,* 1947, **61,** No. 4.

[7] Loevinger, J. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychol. Bull.,* 1948, **45,** 507-529.

[8] White, B. W. and Saltz, E. Measurement of reproducibility. *Psychol. Bull.,* 1957, **54,** 82-99.

Manuscript received 12/18/57

Revised manuscript received 6/16/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

SIMPLE STATISTICAL METHODS FOR SCALOGRAM ANALYSIS*

LEO A. GOODMAN†

UNIVERSITY OF CHICAGO

Simple statistical methods are developed to test whether coefficients of reproducibility, of homogeneity, or of consistency differ significantly from what can be expected if responses to different items are statistically independent. Simple methods are also developed for estimating the variance of coefficients of reproducibility when it is not assumed that responses to different items are independent. These estimates are used to test whether a coefficient differs significantly from any prescribed value, and also to obtain confidence intervals for these coefficients. The rationale for the measurement of reproducibility is also discussed.

The desirability and importance of developing statistical methods in scalogram analysis for the comparison of an observed coefficient (of reproducibility) with its expected value, estimated under the assumption that responses to different items are statistically independent, has been pointed out in [4] and [7]. In the following three sections such statistical methods are developed for the following coefficients: (*a*) Loevinger's coefficient of homogeneity [10], (*b*) Green's coefficient of consistency [6], (*c*) a coefficient of reproducibility given by Green [6], and (*d*) a coefficient of reproducibility given by Sagi [13]. These methods can be used to test whether each of these coefficients is significantly different from what can be expected if responses to different items are statistically independent. Of course, these methods are to be applied to a given coefficient only in situations where it seems appropriate to use that coefficient [cf. 6, 10, 11, 13]. These methods can be used to determine "whether a set of items is worth purifying and scaling" [cf. 13], and should not be applied, except with great caution, to a set of items that have been purified.

Although the comparison of an observed coefficient with its expected value, estimated under the assumption that responses to different items are independent, is useful for some purposes, this particular kind of comparison may be neither very informative nor useful for other purposes. It may be more or less obvious that the responses are not independent and that a coefficient does differ significantly from what can be expected if independence were assumed. It may be more worth while to compare the observed data with some model other than that obtained by the assumption of independence. The problem of testing whether a coefficient differs significantly from what can be expected under the assumption of independence may not be identical with the problem of testing whether the observed data indicate that there is, in some sense, a scale. Also, the problem of measuring the size of the difference between an observed coefficient and what can be expected in the case of independence may not be identical with the problem of measuring, in some sense, how scalable the observed data are. In the present paper these problems are discussed briefly, and methods are presented for estimating the variance of Guttman's coefficient, or of a number of other coefficients, of reproducibility [cf. 13] when the assumption of independence is not made. These methods lead to the development of tests of whether an observed coefficient of reproducibility is significantly different from any prescribed value. Also described are methods of obtaining confidence intervals for coefficients of reproducibility that can be used for the kinds of situations discussed in [6, 10, 13]. For the sake of completeness, the situation in which there is a random sample of items also is considered.

---

*This research was carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research, and of the Social Science Research Committee, University of Chicago. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author is indebted to Jacob Gewirtz for some very helpful comments.

†This paper is dedicated to Professor Jacob Marschak on the occasion of his sixtieth birthday, July 23, 1958.

Finally, a numerical comparison is presented between the method suggested in the following section for testing whether Guttman's coefficient differs significantly from what can be expected if independence is assumed and the method suggested in [13] for performing this test. One of the serious limitations of the method suggested in [13], which was mentioned by the author, is that "when N (the number of respondents in the sample) is greater than 10, and K (the number of items) is 4 or greater, the labor of computing ··· becomes excessive." The method described here is very much simpler to use than that given in [13], and it will be appropriate when the number of respondents in the sample (i.e., the sample size) N is large. It can be used without any real difficulty, for any number K (either small, moderate, or large) of items.

The methods presented in the present paper are approximations in the sense that they depend on certain asymptotic results, i.e., as N approaches infinity the methods become less approximate. Many other statistical methods are also approximations in this same sense. The numerical comparisons suggest that, even when the sample size is quite small, the methods presented lead to rather good approximations in some particular cases. A modification of these methods, which leads to still better approximations for small samples, is also suggested. It would be valuable to compare the various approximate methods presented here in many other particular cases in order to determine how large a sample must be drawn for these methods to lead to useful approximations; this would be worth while with regard to the approximate methods presented here as well as to many other statistical methods (cf. [2], p. 328).

A Coefficient of Reproducibility

The coefficients given in [6, 11, and 13] are all defined for an ordered set of K items in which there are only two kinds of responses to each item. In these articles it was assumed that the ordering of the K items can be obtained from a priori knowledge [cf. 11, 13], or be determined directly by arranging "the items in rank order according to their popularities" [6]. It also will be assumed here either that the ordering of the items is known a priori, or that the total number of respondents sampled is sufficiently large so that almost certainly the rank order according to popularities will be the same as that obtained if the total population of respondents were studied. In the latter case, it is assumed that the rank ordering according to popularities obtained for the total population is unambiguous (i.e., that the "popularities" of two different items will not be exactly equal for the total population) and that this rank ordering corresponds to the "true" ordering of the K items. The situation where there are more than two kinds of responses to each item was commented upon in [6, 7, and 9], and it will not be discussed here.

For $K = 5$, the response pattern $(+-++-)$ indicates that the respondent gives positive responses to items 1, 3, 4, and a negative response to items 2, 5. The following response patterns would be ideal, indicating a "perfect scale" [cf. 4]: $(-----)$, $(----+)$, $(---++)$, $(--+++)$, $(-++++)$, $(+++++)$. An observed response pattern may differ from the ideal patterns; there are several ways of measuring this "error."

In [13], the total error e' is obtained by comparing item i with item $i + 1$, counting the number of $(+-)$ patterns observed for these adjacent items (i.e., the number $f_{i,i+1}$ of individuals who answered positively on item i and negatively on item $i + 1$), and summing these counts for all i; i.e.,

$$e' = \sum_{i=1}^{K-1} f_{i,i+1} \,.$$

For $K \leq 3$, this method of computing the total error is identical with Guttman's method [cf. 13]. Since the coefficient of reproducibility is based directly on the observed total error, i.e., $c' = 1 - (e'/NK)$, a method for testing whether the observed e' differs significantly from what can be expected when the responses to different items are independent will be given first. If the number of individuals who answered item i positively is f_i, and the number of negative responses on item i is $N - f_i = \bar{f}_i$, then the expected value (given f_i and f_{i+1}) of $f_{i,i+1}$ under the assumption of independence, is $f_i\bar{f}_{i+1}/N$. Also if the assumption of independence is true, using the usual approach to the problem of testing the null hypothesis of independence in a 2×2 contingency table (or the null hypothesis that the expected value of the difference between two proportions, obtained from independent samples, is zero), it can be seen [cf. 12] that the distribution of the statistic $[f_{i,i+1} - (f_i\bar{f}_{i+1}/N)]^2/$

$(f_i\bar{f}_i f_{i+1}\bar{f}_{i+1}/N^3)$ will be approximately a χ^2 distribution with one degree of freedom, as N approaches infinity. Furthermore, the distribution of the statistic $[f_{i,i+1} - (f_i\bar{f}_{i+1}/N)]/(f_i\bar{f}_i f_{i+1}\bar{f}_{i+1}/N^3)^{1/2}$ will be approximately a normal distribution with zero mean and unit variance, if the assumption of independence is true, and $N \to \infty$ (cf. [12], p. 71). It can also be seen that, if the assumption of independence is true, the distribution of the statistic

$$\frac{\sum_{i=1}^{K-1} [f_{i,i+1} - (f_i\bar{f}_{i+1}/N)]}{\left(\sum_{i=1}^{K-1} f_i\bar{f}_i f_{i+1}\bar{f}_{i+1}/N^3\right)^{1/2}}$$

will be approximately a normal distribution with zero mean and unit variance (see Appendix). Thus, the null hypothesis of independence of response can be tested by seeing whether e' differs significantly from

$$E' = \sum_{i=1}^{K-1} f_i\bar{f}_{i+1}/N;$$

i.e., by computing $(e' - E')/S_{e'} = z'$, where

$$S_{e'}^2 = \sum_{i=1}^{K-1} f_i\bar{f}_i f_{i+1}\bar{f}_{i+1}/N^3,$$

and comparing the numerical value of z' with the standard unit normal distribution. For example, an approximate one-sided test, at the 95 per cent level of significance, can be obtained by rejecting the null hypothesis whenever $z' \leq -1.645$.

Since $c' = 1 - (e'/NK)$, $z' = -(c' - C')/S_{c'}$, where $C' = 1 - (E'/NK)$ and $S_{c'} = S_{e'}/NK$. Thus, an approximate one-sided test, at the 95 per cent level of significance, can be obtained if the null hypothesis is rejected whenever $c' > C' + (1.645)S_{c'}$.

A Coefficient of Homogeneity

Rather than compute e' as was done above, consider now the total error e'' obtained by comparing item i with item j $(i < j)$, counting the number of $(+-)$ patterns observed for these two items (i.e., the number $f_{i,j}$ of individuals who responded positively on item i and negatively on item j), and summing these counts for all i and j $(i < j)$; i.e.,

$$e'' = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} f_{i,j} .$$

The expected value (given the f_i) of e'', when responses to different items are independent, is

$$E'' = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} f_i\bar{f}_j/N.$$

Loevinger suggests $(E'' - e'')/E''$ as a coefficient h of homogeneity [cf. 10, 15]. A simple test of the null hypothesis that responses to different items are independent is based on determining whether h differs significantly from zero.

By an approach similar to that presented above (cf. Appendix), it is seen that, if the null hypothesis is true, then the distribution of the statistic $(E'' - e'')/S_{e''} = z''$ will be approximately a normal distribution with zero mean and unit variance, when $N \to \infty$, and

$$S_{e''}^2 = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} (f_i \bar{f}_i f_j \bar{f}_j / N^3).$$

Thus, an approximate one-sided test, at level of significance .95, can be obtained by rejecting the null hypothesis whenever $z'' > 1.645$. Also, since h is such that $hE''/S_{e''} = z''$, an approximate test of the null hypotheses at level of significance .95 can be obtained if this hypothesis is rejected whenever $h > 1.645\ S_{e''}/E''$. (For this hypothesis, the expected value of h is zero.)

The method presented here is an approximation in the same sense as is the method presented in the earlier section. The results of some numerical comparisons presented in the final section tend to suggest that the approximate test of whether h differs significantly from zero will be quite useful. The modification described in the final section to improve the approximation when sample sizes are small or moderate can also be applied to obtain from $S_{e''}$ a modified statistic $S_{e''}^*$.

Another Coefficient of Reproducibility

Now consider one of the coefficients Rep_B in [6], suggested as an estimate of reproducibility. (The author states that this coefficient was checked empirically, and that the average discrepancy between this coefficient and the "correct" coefficient of reproducibility was .003 [cf. 6].) For this coefficient, the total error e''' is obtained as

$$e''' = \sum_{i=1}^{K-1} f_{i,i+1} + \sum_{i=2}^{K-2} (f_{i,i+2} f_{i-1,i+1})/N,$$

and $c''' = 1 - (e'''/NK)$. The expected value (given the f_i) of e''', under the assumption that responses to different items are independent, is

$$E''' = \frac{\sum_{i=1}^{K-1} f_i \bar{f}_{i+1}}{N} + \frac{\sum_{i=2}^{K-2} f_i \bar{f}_{i+2} f_{i-1} \bar{f}_{i+1}}{N^3}.$$

A simple test of the null hypothesis that responses are independent is to determine whether e''' differs significantly from E'''.

By an approach similar to that presented above, it can be seen that, if the hypothesis of independence is true, then the distribution of the statistic $(E''' - e''')/S_{e'''} = z'''$ will be approximately normal in form with zero mean and unit variance when $N \rightarrow \infty$, and

$$S^2_{e'''} = S^2_{e'} + \left\{ \sum_{i=2}^{K-3} [S^2_{i,i+2}(E_{i-1,i+1} + E_{i+1,i+3})^2] + S^2_{1,3}E^2_{2,4} + S^2_{K-2,K}E^2_{K-3,K-1} \right\} / N^2,$$

where

$$E_{i,j} = f_i\bar{f}_j/N \quad \text{and} \quad S^2_{i,j} = f_i\bar{f}_i f_j\bar{f}_j/N^3.$$

(This result is derived in the Appendix.) Thus, an approximate one-sided test, at level of significance .95, can be obtained by rejecting the null hypothesis whenever $z''' > 1.645$. Also, since $c''' = 1 - (e'''/NK)$, $z''' = (c''' - C''')/S_{c'''}$, where $C''' = 1 - (E'''/NK)$, and $S_{c'''} = S_{e'''}/NK$. Hence, an approximate one-sided test, at the level of significance .95, can be obtained if the null hypothesis is rejected whenever $c''' > C''' + (1.645)S_{c'''}$.

Green [6] also suggests an index of consistency $g = (c''' - C''')/(1 - C''')$. Since $g(1 - C''')/S_{c'''} = z'''$, an approximate test of the hypothesis of independence at level of significance .95 can be obtained if this hypothesis is rejected whenever $g > 1.645\ S_{c'''}/(1 - C''')$. (For this hypothesis, the expected value of g is zero.)

Some Different Problems and Methods

In the preceding sections, all of the formulas for the expected values (the E's) and the S^2's were computed under the assumption that responses to different items are statistically independent, and the large sample distribution theory results also were obtained under this assumption. Now this assumption will not be made, and a somewhat different problem will be studied. The methods presented in this section can be applied to Guttman's coefficient of reproducibility, as well as to a number of other coefficients, and to the problem of testing whether the expected value of a given coefficient is some prescribed numerical value. Also, these methods can be used to obtain approximate confidence intervals for the expected value of the coefficient.

Consider the situation where there is an ordered set of items, with only two kinds of responses to each item. Assume that a sample of respondents was drawn at random from an infinite population or, for a finite population, that sampling was done with replacement. Since there are 2^K different possible patterns of responses, the total population could conceivably be divided up into 2^K different categories corresponding to these patterns. To each category i.e., to each pattern of responses, a number is assigned indicating how different that pattern is from an ideal one; e.g., to pattern $(++---)$ the number 1 is assigned for the method in [13], and the number 2 is assigned for Guttman's method [cf. 13]. The latter method is referred to as Guttman's method in [13]—this title is also used here although it may differ somewhat from the methods in [7] and [8]. The different methods lead to different coefficients of reproducibility. The number assigned to an ideal pattern, e.g., $(---++)$, is zero. Hence, for any given method of assigning numbers, the tth pattern, $t = 1, 2, \cdots, 2^K$, will be assigned a number X_t, and the population coefficient of reproducibility is

$$C = 1 - \frac{\sum_t D_t X_t}{\bar{N}K},$$

where $D_t/\bar{N} = P_t$ is the proportion of the population in category t. The coefficient in [13] and also the coefficient obtained using Guttman's method are both of this general form; Green's coefficient Rep_B, i.e., c''', differs somewhat from this form, but his coefficient Rep_A is also of this general form [cf. 6, 13]. All of the coefficients mentioned here of this general form have the property $X_t \leq K/2$ for all t; e.g., the maximum value of X_t, for these coefficients, occurs when the pattern is of the kind $(+-+-+-)$, where $X_t = K/2$ (for K even). Thus, $Y_t = X_t/K \leq 1/2$, and

$$C = 1 - \frac{\sum_t D_t Y_t}{\bar{N}} = 1 - \sum_t P_t Y_t .$$

The form of the coefficient W, presented by Jardine in [9], is quite different from the general form described here. The coefficient W measures a different aspect of the data, although this coefficient can be used to develop still another test of the null hypothesis discussed in the preceding sections, i.e., that responses to the different items are statistically independent [cf. 9]. Jardine mentions that "a test for W in the non-null case is not available." In the present section, tests in the "non-null case" for coefficients of the general form described above are presented.

If C is understood to be a measure of scalability (or a definition of the amount of scalability), then it is clear that the system of numbers X_t (or the Y_t) is a basic part of this definition. For different systems of Y_t, different definitions of scalability are obtained. Which system (i.e., which measure or definition) will be appropriate depends on the particular context to which the measure is to be applied (cf. discussion of measures of association in [5]). For example, in some context C^*, based on using $Y_t = 1/2$ for all categories t that do not correspond to ideal patterns, will be appropriate. This measure is closely related to the proportion of the population falling into all categories that correspond to ideal patterns, and C^* will be useful if the context is such that the important factor is this proportion. If prediction of a response pattern for an individual is to be based on his total score, and on the ideal pattern corresponding to that total score for the set of items, then the proportion of correct predictions (of response patterns) for the population will be $2C^* - 1$. Thus, for some purposes $2C^* - 1$ (or C^*) measures an aspect of the reproducibility of the response patterns from the total scores and the set of ideal patterns.

In some contexts, it may be preferable not to use the collection of $K + 1$ ideal patterns as the possible predicted patterns, but rather some other collection of patterns as the predicted ones. Thus, sometimes the proportion of correct predictions of response patterns for the population can be increased by using the $K + 1$ patterns that occur most frequently in the population as the collection of possible predicted patterns. Also, for some purposes, this collection need not consist of $K + 1$ patterns; perhaps some larger number would be worth while. Reproducibility in the sense described above can sometimes be increased if the number of possible predicted patterns is increased. Hence, reproducibility in this sense can be high in situations where C^* or some other measure of scalability is low. It might have been preferable to use the term coefficient of scalability for the coefficients of reproducibility given in [6, 7, and 13]. However, the term reproducibility does suggest an operational concept that is worthy of emphasis [cf. 5]. In some contexts, the proportion of item responses, rather than the proportion of response patterns predicted correctly, will be important. For these situations C^* would not be an appropriate measure, but some other coefficients might be [cf. 7, 8, 15].

An estimate of the coefficient of reproducibility in the case where a random sample of N respondents is observed, would be

$$c = 1 - \frac{\sum_t d_t X_t}{NK} = 1 - \frac{\sum_t d_t Y_t}{N},$$

where d_t/N is the proportion of the sample in category t. The expected value of c is C, and the variance of c is

$$\sigma_c^2 = \frac{\sum_t P_t Y_t^2 - (\sum_t P_t Y_t)^2}{N}.$$

Since $Y_t \leq 1/2$,

$$\sigma_c^2 \leq \frac{\sum_t P_t Y_t/2 - (\sum_t P_t Y_t)^2}{N} = \frac{(1 - C)/2 - (1 - C)^2}{N}$$

$$= \frac{(1 - C)(C - \frac{1}{2})}{N} = \bar{\sigma}^2.$$

This gives an upper bound $\bar{\sigma}^2$ for the exact variance σ_c^2 of the coefficient of reproducibility. For some coefficients (i.e., for some assignments of numbers Y_t), the exact variance σ_c^2 will be very close to the upper bound $\bar{\sigma}^2$ (e.g., for c^*, σ_c^2 will be equal to $\bar{\sigma}^2$). For other coefficients, the bound $\bar{\sigma}^2$ may be much larger than σ_c^2. For any coefficient, σ_c^2 will be equal to $\bar{\sigma}^2$ if $P_t = 0$ for all categories t where the assigned numbers Y_t differ from 0 or 1/2. When $\sigma_c^2 > 0$, using the usual central limit theorem, the distribution of $z = (c - C)/\sigma_c$ will be approximately normal with zero mean and unit variance, when $N \to \infty$. Also $\hat{z} = (c - C)/\hat{\sigma}_c$, where

$$\hat{\sigma}_c^2 = \frac{\sum_t d_t Y_t^2 - [(\sum_t d_t Y_t)^2/N]}{N(N-1)},$$

will be approximately normal with zero mean and unit variance. Furthermore, $\bar{z} = (c - C)/\bar{\sigma}$ and $\hat{\bar{z}} = (c - C)/\hat{\bar{\sigma}}$, where $\hat{\bar{\sigma}}^2 = (1 - c)(c - 1/2)/N$, will be approximately normal with zero mean and variance ≤ 1. Thus, if the null hypothesis is that C is a specified value, then an approximate one-sided test, at level of significance .95, can be obtained by rejecting the null hypothesis whenever $c > C + (1.645)\,\hat{\sigma}_c$. A simpler approximate (and conservative) one-sided test, at level of significance $\leq .95$, can be obtained by rejecting the null hypothesis whenever $c > C + (1.645)\bar{\sigma}$. For some purposes, a two-sided test may be appropriate; the general method described herein also leads directly to these two-sided tests. Using this method, it is also possible to obtain one or two-sided confidence intervals for C. Thus the interval $c - (1.96)\,\hat{\sigma}_c < C < c + (1.96)\,\hat{\sigma}_c$ is an approximate 95 per cent confidence interval for C. Also, the interval $c - (1.96)\,\hat{\bar{\sigma}} < C < c + (1.96)\,\hat{\bar{\sigma}}$ is an approximate (conservative) confidence interval at a level $\leq .95$.

The preceding comments pertain to the situation where a random sample of individuals is drawn, and the numerical value of X_t can be determined for tth response pattern, $t = 1, 2, \cdots, 2^K$. For a fixed, i.e., given, response pattern t, the number X_t is determined by the definition of scalability used, and this number is considered fixed. In some situations, it may be worth while to consider this number as a random variable x_t rather than as a fixed value X_t. This will be the case if X_t is the number assigned to the tth category defined for a given population of K items, and x_t is the number obtained for this category when a random sample of k items is drawn from this population of K items ($k < K$), and the response patterns to these sample items are used to compute x_t. The case where there is a random sample of items will be discussed for the sake of completeness, although in many situations it may not be appropriate to consider the actual set of items used as a random sample from a larger population.

In this case, if the coefficient is computed on the basis of the observed response patterns to the k sample items, then this coefficient will in general differ from the corresponding coefficient for the population of K items. In the special situation where the K items form a perfect scale, the coefficients would be identical. In fact, the expected value of the coefficient based on k items, for all possible samples of k items, will generally differ from the coefficient based on K items. For example, in the extreme case where a population consists of $K = 4$ items, and where $(+-+-)$ is the only response pattern, then the population value of $2C^* - 1 = 0$, but this measure computed from a random sample of two items drawn from the 4 will take on the values 0 or 1, depending on which two items are drawn. The expected value of this measure will be .5 when items are sampled without replacement. In this extreme case, C', C^*, and Guttman's coefficient will all take on the values .5 or 1 depending on which two items are drawn; the expected value of these coefficients will be .75, while the population coefficient is .5. Thus, the coefficients of reproducibility computed for a sample of items from a larger population of items must be interpreted with great caution. They generally do not lead to unbiased estimates of the corresponding coefficients for the larger population, and they can give quite misleading results.

Now consider the set of K ordered items that are used in an empirical study as a random set of ordered items; i.e., there is a population of M sets, each containing K ordered items, and one of these sets of items is randomly selected for the empirical study. Also consider the situation where the responses of a given individual j to the M sets of items are obtained, and $1 - \sum_t {}_jR_tY_t = C_j$ is computed for the population of sets of items for this individual j, where ${}_jR_t$ is the proportion of sets among the M sets where the response pattern for individual j corresponds to the tth category, for $t = 1, 2, \cdots, 2^K$. Thus, in an empirical study where a single set of K items is used, the numerical value $c_j = 1 - {}_jY$ corresponding to the jth individual in the population, where ${}_jY = {}_jX/K$ is the number associated with the particular response pattern observed for the jth individual can be considered an estimate of C_j for that individual.

This c_j will be an unbiased estimate of C_j and the variance of this estimate is $\sum_t {}_jR_tY_t^2 - (\sum_t {}_jR_tY_t)^2 = {}_j\sigma_Y^2$. If c_j is computed for each individual in a given population of $\bar{N}$ individuals, then $\sum_j c_j/\bar{N}$ will be equal to the coefficient of reproducibility c computed for a single set of K items for a population of $\bar{N}$ individuals. Also, the expected value of c is $\sum_j C_j/\bar{N} = \bar{C}$, which would be the coefficient of reproducibility computed on the basis of all M sets of K items and all individuals in the population. The variance of c is

$$\sigma_M^2 = \sum_{m=1}^{M} (c^{(m)} - \bar{C})^2/M,$$

where $c^{(m)}$ denotes the coefficient of reproducibility computed for the mth set $(m = 1, 2, \cdots, M)$ of ordered items in the population of M sets of items. In the special situation in which, for each respondent in the population, a set of items is drawn at random from the population of sets (or if the probability of response pattern t for individual j is ${}_jR_t$, and this is statistically independent of the response pattern obtained for a different individual), then the variance of c will be somewhat different; viz.,

$$\sigma_M^{*2} = \sum {}_j\sigma_Y^2/\bar{N}^2 = \bar{\sigma}_Y^2/\bar{N},$$

where $\bar{\sigma}_Y^2$ is the average value of ${}_j\sigma_Y^2$ for the total population.

For the case in which a random sample of N respondents is drawn from a large population the coefficient of reproducibility c computed for the sample responses to a single set of items will be an unbiased estimate of $\bar{C}$; the variance of c will be $\bar{\sigma}_M^2 + (\sigma_C^2/N)$, where σ_C^2 is the variance of C_j for the total population of respondents, $\bar{\sigma}_M^2$ is the average of

$$\sum_{m=1}^{M} (c_{(s)}^{(m)} - \bar{C}_{(s)})^2/M = {}_{(s)}\sigma_M^2$$

for all samples ($s = 1, 2, \cdots$) of size N, and $c_{(s)}^{(m)}$ and $\bar{C}_{(s)}$ are the corresponding values of $c^{(m)}$ and $\bar{C}$ computed for the sth sample, by considering this sample as a population of size N. Also, in the special case where the probability of response pattern t for individual j is ${}_jR_t$, and this probability is statistically independent of the response pattern for a different individual, then the variance of c will be $(\bar{\sigma}_Y^2 + \sigma_C^2)/N$. In this case, if it can be assumed also that ${}_jR_t = R_t$, thus, $C_j = C$ and ${}_j\sigma_Y^2 = \sigma_Y^2$ for all individuals $j = 1, 2, \cdots$, then the expected value of the coefficient of reproducibility c, computed for a single set of K items for a random sample of N individuals, is C; the variance of c is $\sigma_Y^2/N = \sigma_c^2$. The latter was computed earlier in this section; an upper bound for σ_c^2 was also presented.

Some Numerical Comparisons

In [13] the exact distribution of e' is presented for some special cases. Now the results obtained using these exact distributions will be compared with those obtained using the approximate methods presented in earlier sections.

As seen in Table 4 in [13], if $N = 100$, $f_1 = 20$, $f_2 = 40$, $f_3 = 60$, $f_4 = 80$, then the null hypothesis that responses to different items are statistically independent would be rejected at the level of significance $\geq .95$ if $e' \leq 33$. At the level of significance $\geq .975$, rejection of the null hypothesis would occur if $e' \leq 32$. At the level of significance $\geq .99$, rejection would occur if $e' \leq 30$.

The approximate method described in an earlier section indicates that, at level of significance $\geq .95$, the null hypothesis should be rejected if $z' \leq 1.645$; i.e., if $e' \leq 40 - (1.645)\ (3.67) = 33.96$. Thus, in this case, the approximate test and the exact test lead to the same critical region, $e' = 0, 1, 2, \cdots, 33$. At level of significance $\geq .975$, rejection would occur if $e' \leq 40 - (1.960)\ (3.67) = 32.81$, i.e., if $e' \leq 32$; at level of significance $\geq .99$, rejection would occur if $e' \leq 40 - (2.326)\ (3.67) = 31.46$, i.e., if $e' \leq 31$.

From the preceding numerical computations, it is noted that the only difference between the exact method and the simpler approximate method is obtained at the level of significance $\geq .99$. Actually, this difference at the

$\geq .99$ level is due in part to the fact that the exact level of significance of the critical region $e' \leq 30$ is $1 - .00495 = .995$, while the exact level of significance of the critical region $e' \leq 31$ is $1 - .011051 = .989$ [cf. 13]. Thus, the approximate method leads to quite an accurate approximation in the particular case considered even at level of significance $\geq .99$. In this case, the method is conservative in the sense that it will lead to rejection of the null hypothesis whenever the exact method would have led to rejection, but it may also lead to rejection in some cases where the exact method would not (cf. [14], p. 105).

From Table 3 in [13], if $N = 10$, $f_i = 5$ for $i = 1, 2, 3, 4$, then the null hypothesis would be rejected, at level of significance $\geq .95$, if $e' \leq 4$. At level of significance $\geq .99$, rejection would occur if $e' \leq 3$. The approximate method leads to rejection, at level of significance $\geq .95$, if $e' \leq 7.5 - (1.645)(1.37) = 5.24$, i.e., $e' \leq 5$; also, at level of significance $\geq .99$, the approximate method leads to rejection if $e' \leq 7.5 - (2.326)(1.37) = 4.31$, i.e., $e' \leq 4$. Thus, in this particular case where $N = 10$, the critical regions obtained by this method differed by one unit from the critical regions obtained using the exact distribution. Here, too, differences between the results obtained by the approximate and exact methods are due, in part, to the discontinuous nature of the exact distribution (cf. [2], p. 331). In this case, too, the approximate method leads to a conservative test.

For medium and rather small samples, a modification of the approximate methods (changing an N to an $N - 1$) may lead to improved approximations. This modification is based on the fact that, when the null hypothesis is true, the statistic

$$\frac{f_{i.i+1} - (f_{i}\bar{f}_{i+1}/N)}{\{f_{i}f_{i}f_{i+1}\bar{f}_{i+1}/[N^2(N-1)]\}^{1/2}}$$

will be approximately a unit normal variate (cf. [12], p. 71), and $(e' - E')/S_{e'}^*$ will also be approximately a unit normal, where

$$S_{e'}^* = \left[\sum_{i=1}^{K-1} \{f_{i}\bar{f}_{i}f_{i+1}\bar{f}_{i+1}/[N^2(N-1)]\}\right]^{1/2}.$$

Thus, for the preceding example, $S_{e'}^* = [3(625)/(1000\text{–}100)]^{1/2} = 1.44$, and this modified method leads to rejection of the null hypothesis, at level of significance $\geq .95$, if $e' \leq 7.5 - (1.645)(1.44) = 5.13$, i.e., $e' \leq 5$. Also, at level of significance $\geq .99$, this method leads to rejection if $e' \leq 7.5 - (2.326)(1.44) = 4.15$, i.e., $e' \leq 4$. Thus, a slight improvement is obtained, although it is not sufficient to change the critical regions given earlier using the $S_{e'}$ rather than $S_{e'}^*$.

Since the approximate methods given here are based, in part, on the same kinds of asymptotic results as the usual approach to the problem of testing independence in a 2×2 contingency table, the reader is referred to

[2] for discussion and recommendations concerning the 2 × 2 contingency table and also for some references to tables of exact distributions that can be of some use in computing the exact distribution of e' when N and K are small. The reader will note that the modification presented in this section to improve the approximate methods is not related to the usual Yates correction (cf. [2], p. 332). A modification related to the Yates correction is not recommended because, for the 2 × 2 contingency table, this correction "has a tendency to over-correct" and this "over-correction mounts up" [2], when the corrected statistics for each of a number of 2 × 2 tables are combined.

It has been shown that the approximate method for testing the hypothesis that responses to different items are independent, and its modified form given here, lead to quite useful results in two cases. If N is sufficiently large to permit the application of the usual χ^2 test for independence to each of the $(K - 1)$ contingency tables obtained by comparing responses for items i and $i + 1$ $(i = 1, 2, \cdots, K - 1)$, then the approximate method, or its modified form, leads to useful results.

Appendix

When the responses to different items are statistically independent, the statistic $(E''' - e''')/S_{e'''} = z'''$, can be shown to be approximately normally distributed with zero mean and unit variance when $N \to \infty$. A similar approach can be used to prove corresponding results for the statistics z'' and z'.

The statistic $e''' - E'''$ can be written as

$$e''' - E''' = \sum_{i=1}^{K-1} (f_{i,i+1} - E_{i,i+1}) + \sum_{i=2}^{K-2} (f_{i,i+2}f_{i-1,i+1} - E_{i,i+2}E_{i-1,i+1})/N$$

$$= \sum_{i=1}^{K-1} \Delta_{i,i+1} + \sum_{i=2}^{K-2} (f_{i,i+2}\Delta_{i-1,i+1} + E_{i-1,i+1}\Delta_{i,i+2})/N,$$

where $\Delta_{i,j} = f_{i,j} - E_{i,j}$. The expected value (given the numbers f_i) of $e''' - E'''$ will be zero, since the expected value of $\Delta_{i,j}$ will be zero, and the statistics $f_{i,i+2}$ and $\Delta_{i-1,i+1}$ will be statistically independent (given the f_i). $(e''' - E''')N^{-1/2} = w$ may also be written as

$$N^{-1/2}\left\{\sum_{i=1}^{K-1} \Delta_{i,i+1} + \sum_{i=2}^{K-2} (E_{i,i+2}\Delta_{i-1,i+1} + E_{i-1,i+1}\Delta_{i,i+2} \right.$$
$$\left. + \Delta_{i,i+2}\Delta_{i-1,i+1})/N\right\}$$

Each $\Delta_{i,j}N^{-1/2} = v_{ij}$ will asymptotically normally distributed with zero mean and an asymptotic variance (for fixed number f_i/N) of $S_{ij}^2 = f_i\bar{f}_if_j\bar{f}_j/N^4$ (cf. [12], p. 71; [3], p. 366). Thus,

$$w = \sum_{i=1}^{K-1} v_{i,i+1} + \sum_{i=2}^{K-2} (E_{i,i+2}v_{i-1,i+1} + E_{i-1,i+1}v_{i,i+2} + v_{i,i+2}v_{i-1,i+1}N^{1/2})/N,$$

$$= \sum_{i=1}^{K-1} v_{i,i+1} + [v_{1,3}E_{2,4} + v_{K-2,K}E_{K-3,K-1} + \sum_{i=2}^{K-3} v_{i,i+2}(E_{i-1,i+1} + E_{i+1,i+3})]/(N + \delta),$$

where

$$\delta = \sum_{i=2}^{K-2} v_{i,i+2}v_{i-1,i+1}N^{-1/2}.$$

Since δ will converge in probability to zero, the statistic w is essentially a linear combination of the v_{ij}'s (cf. [3], p. 254). The v_{ij}'s are asymptotically independent—this follows from the assumption of independence of the responses, or it could also be proved using the methods given in [1]. Hence the statistic w is essentially a linear combination of asymptotically normally distributed independent random variables, and this statistic will also be asymptotically normally distributed. The asymptotic variance (for fixed f_i/N) of this statistic will be

$$S^2_{e'''} = \sum_{i=1}^{K-1} S^2_{i,i+1} + \frac{S^2_{1,3}E^2_{2,4} + S^2_{K-2,K}E^2_{K-3,K-1} + \sum_{i=2}^{K-3} S^2_{i,i+2}(E_{i-1,i+1} + E_{i+1,i+3})^2}{N^2}.$$

REFERENCES

[1] Anderson, T. W. and Goodman, L. A. Statistical inference about Markov chains. *Ann. math. Statist.*, 1957, **28**, 89-110.

[2] Cochran, W. G. The χ^2 test of goodness of fit. *Ann. math. Statist.*, 1952, **23**, 315-345.

[3] Cramer, H. *Mathematical statistics.* Princeton: Princeton Univ. Press, 1946.

[4] Festinger, L. The treatment of qualitative data by scale analysis. *Psychol. Bull.*, 1947, **44**, 149-161.

[5] Goodman, L. A. and Kruskal, W. H. Measures of association for cross classifications. *J. Amer. statist. Ass.*, 1954, **49**, 732-764.

[6] Green, B. F. A method of scalogram analysis using summary statistics. *Psychometrika*, 1956, **21**, 179-188.

[7] Guttman, L. The basis for scalogram analysis. In S. A. Stouffer et al., *Measurement and prediction.* Princeton: Princeton Univ. Press, 1950. Pp. 60-90.

[8] Guttman, L. On Festinger's evaluation of scale analysis. *Psychol. Bull.*, 1947, **44**, 451-465.

[9] Jardine, R. Ranking methods and the measurement of attitudes. *J. Amer. statist. Ass.*, 1958, **53**, 720–728.

[10] Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, **61**, No. 4.

[11] Loevinger, J. The technic of homogeneous tests compared with some aspects of "scale analysis." *Psychol. Bull.*, 1948, **45**, 507-529.

[12] Pearson, E. S. and Hartley, H. O. *Biometrika tables for statisticians*, Vol. 1. Cambridge: Cambridge Univ. Press, 1954.

[13] Sagi, P. C. A statistical test for the significance of the coefficient of reproducibility. *Psychometrika*, 1959, **24**, 19-27.

[14] Walker, H. M. and Lev, J. *Statistical inference*. New York: Holt, 1953.

[15] White, B. W. and Saltz, E. Measurement of reproducibility. *Psychol. Bull.*, 1957, **54**, 81-99.

Manuscript received 4/3/58

Revised manuscript received 6/11/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

HIERARCHICAL FACTOR SOLUTIONS WITHOUT ROTATION

ROBERT J. WHERRY

OHIO STATE UNIVERSITY

A method is presented for securing a hierarchical factor solution which achieves simple structure at each hierarchical level without rotation or even preliminary arbitrary orthogonal or oblique solutions. The method is based upon the assumption that if overlap is removed from clusters the remaining specifics will achieve simple structure automatically. The problem presented earlier by Schmid and Leiman, using oblique simple structural rotation as a basis, is reworked by this new approach.

Schmid and Leiman [1] recently published an article on the development of hierarchical factor solutions. Their method was an extension of a method proposed by Thomson [2] a number of years ago. It consisted of factoring and rotating the factors at the first-order level to oblique simple structure. The factor correlation matrix was then obtained by inverting the reference vector correlation matrix and normalizing it. This factor correlation matrix was in turn factored and the process was repeated at this higher level. This cycle was repeated until a single factor explained the highest level factor matrix. A method of back-solution was then provided which yielded a transformation matrix to be applied to the next lower level oblique simple structure loadings for each order involved.

The author, with the assistance of Dr. B. J. Winer, had worked out this identical extension independently. The method was still considered too cumbersome and indirect, however, and the search for a more direct solution continued. With suggestions from several sources, particularly from Dr. Roderick Bare, the author succeeded in eliminating the need for oblique cluster loadings, their transformation to orthogonality, the subsequent rotation to simple structure, and the need for the reference and factor correlation matrices.

The present method starts with the multiple group factor method, and obtains the usual cluster sums and the cluster correlations. It is assumed that if all overlap is removed from the clusters then they will have simple structure with respect to each other. Hence the multiple group method is applied immediately to the cluster correlation matrix and a new set of higher order cluster sums and cluster correlation matrix is obtained. This process is repeated until the final cluster matrix consists of a single cluster.

It is now assumed that this final cluster represents a general factor with loadings on all of the original tests or variables, and that the specifics at that level represent sub-general factors possessing simple structure with respect to each other and having loadings on all tests contained in the lower order clusters which compose them. Each cluster, since it represents a postulated reference vector, is assumed to have unit variance for purposes of computing the specific loadings; of course, communalities are used in the multiple cluster factorization. The data at this final stage will consist of:

a. the highest level unitary factor cluster matrix, a set of loadings for that factor, and a specific factor loading for each lower order cluster represented in the matrix;
b. one or more lower order analyses for each of which the following are available:
    1. the cluster sums for each still lower order cluster (in the lowest order these are cluster sums for each original variable),
    2. the multipliers obtained by taking the reciprocal of the square root of the sum of cluster sums for the variables in each cluster, which were used to obtain the cluster correlation matrix at that level.

These data are sufficient for obtaining a hierarchical solution identical to that achieved by the Schmid-Leiman approach.

The following steps serve to complete the hierarchical solution.

(1) Using the highest order cluster correlation matrix with ones in the diagonals and the general and specific factor loadings as criteria, obtain the beta weights which would predict these factors from the clusters.

(2) These beta weights are then multiplied by the multipliers referred to in b2 above, and these products are used as a transformation matrix.

(3) These transformation weights are used to multiply the cluster sums of the previous analysis, which yields general and sub-general factor loadings for each of the next lower order clusters.

(4) The communality is computed from these factor loadings for each cluster at this new level, subtracted from unity, and the square root is taken to obtain a specific factor loading for the cluster.

(5) Steps (1) through (4) are repeated for each successive lower order matrix until the original correlation matrix for the variables is involved.

This method has been successfully employed in both actual factoring of correlation matrices and as a modification of the Wherry-Winer [3] method for factoring large numbers of items. It has been used in several studies at Ohio State University and in several industrial concerns. To illustrate the method, it has been applied to the fictitious matrix analyzed by Schmid and Leiman. The solution is presented in Tables 1 through 5.

Table 1 contains the original intercorrelations (R_{00}) and communalities (for this problem assumed to be known). By inspection, each successive pair of variables is seen to constitute a separate cluster (${}_1C_i$), indicated by spacing in the table. Beneath the table are found the six sets of cluster sums (${}_1\sum_i C_i$).

Table 2 gives the sums of the first-order cluster sums (${}_1\sum \sum C_b C_d$), the reciprocals of the square roots of the major diagonal terms (${}_1M_{C'i}$), the

TABLE 1

Intercorrelations of 12 Variables with Communalities in the Diagonals, and Six Sets of Cluster Sums for Each Variable

Cluster	A		B		C		D		E		F	
Variable	1	2	3	4	5	6	7	8	9	10	11	12
	Intercorrelations, R_{oo}											
1	6400	7200	3136	2688	0983	0491	1290	0369	2903	1613	0645	0753
2	7200	8100	3528	3024	1106	0553	1452	0415	3266	1814	0726	0847
3	3136	3528	4900	4200	0753	0377	0988	0282	2222	1235	0494	0576
4	2688	3024	4200	3600	0645	0323	0847	0242	1905	1058	0424	0494
5	0983	1106	0753	0645	6400	3200	1344	0384	1089	0605	0242	0282
6	0491	0553	0377	0323	3200	1600	0672	0192	0544	0302	0121	0141
7	1290	1452	0988	0847	1344	0672	4900	1400	1429	0794	0318	0370
8	0369	0415	0282	0242	0384	0192	1400	0400	0408	0227	0091	0106
9	2903	3266	2222	1905	1089	0544	1429	0408	8100	4500	1458	1701
10	1613	1814	1235	1058	0605	0302	0794	0227	4500	2500	0810	0945
11	0645	0726	0494	0424	0242	0121	0318	0091	1458	0810	3600	4200
12	0753	0847	0576	0494	0282	0141	0370	0106	1701	0945	4200	4900
	Cluster Sums, $_1\Sigma C_{icj}$*											
A	1 3600	1 5300	6664	5712	2089	1044	2742	0784	6169	3427	1371	1600
B	5824	6552	9100	7800	1398	0700	1835	0524	4127	2293	0918	1070
C	1474	1659	1130	0968	9600	4800	2016	0576	1633	0907	0363	0423
D	1659	1867	1270	1089	1728	0864	6300	1800	1837	1021	0409	0476
E	4516	5080	3457	2963	1694	0846	2223	0635	1 2600	7000	2268	2646
F	1398	1573	1070	0918	0524	0262	0688	0197	3159	1755	7800	9100

* Sums for variables in each cluster are underlined.

cluster intercorrelations (R_{11}) with communalities in the diagonals. These communalities were computed by a modified Spearman equation

$$r_{ii}\bar{r}_{ik}/\bar{r}_{jk} \ ,$$

where $\bar{r}_{ik}$ and $\bar{r}_{jk}$ represent the average correlation of variables in the cluster with all variables not in the cluster. Inspection showed this matrix to contain three clusters, indicated by spacing in the table. Beneath the intercorrelation table are the three rows of second-order cluster sums ($_2\sum C_iC_{i'}$).

Table 3 gives the sums of the second-order cluster sums ($_2\sum\sum C_iC_{i'}$), the reciprocals of the square roots of the major diagonal terms ($_2M_{C_i}$), the second-order cluster correlations (R_{22}), with communalities computed as above. This matrix forms a single cluster, and so this table is followed by a row of sums, the actual factor loadings, and the appropriate specific loadings. This concludes the first or forward phase of the solution.

Table 4 shows the details of the second or back-solution phase in some detail. It contains a Doolittle solution using the second-order cluster intercorrelations from Table 3 with ones in the diagonal, and with the loadings from that same table used as successive criteria. The beta weights for each criterion, obtained by the usual back-solution method, are then postmultiplied by a diagonal matrix of multiplier values ($_2M$), also found in Table 3, to obtain a transformation matrix (T_1). Finally the transpose of this trans-

TABLE 2

First-Order Cluster Sum-Sums, Multipliers, and Intercorrelations for Clusters in Table 1, and Second-Order Cluster Sums

Sum-Sums, $_1\Sigma\Sigma C_iD_j$

Cluster	A	B	C	D	E	F
A	2 8900	1 2376	3133	3526	9596	2971
B	1 2376	1 6900	2098	2359	6420	1988
C	3133	2098	1 4400	2592	2540	0786
D	3526	2359	2592	8100	2858	0885
E	9596	6420	2540	2858	1 9600	4914
F	2971	1988	0786	0885	4914	1 6900

Multipliers, $_1M_{ci}$

.588235	.769231	.833333	1.111111	.714286	.769231

Intercorrelations, R_{11}*

Cluster	a		b		c	
A	6400	5600	1536	2304	4032	1344
B	5600	4900	1345	2016	3527	1176
C	1536	1345	1600	2400	1512	0504
D	2304	2016	2400	3600	2268	0756
E	4032	3527	1512	2268	8100	2700
F	1344	1176	0504	0756	2700	0900

Cluster Sums, $_2\Sigma C_{icj}$

a	1 2000	1 0500	2881	4320	7559	2520
b	3840	3361	4000	6000	3780	1260
c	5376	4703	2016	3024	1 0800	3600

* Communalities have been substituted for unities in the diagonals.

TABLE 3

Second-Order Cluster Sum-Sums, Multipliers, and Intercorrelations for Clusters in Table 2, and Third-Order Sums and Factor Loadings

Sum-Sums, $_2\Sigma\Sigma C_iC_j$

Cluster			
a	2 2500	7201	1 0079
b	7201	1 0000	5040
c	1 0079	5040	1 4400

Multipliers, $_2M_{ci}$

666667	1 000000	833333

Intercorrelations, R_{22}

a	6400	4800	5600
b	4800	3600	4200
c	5600	4200	4900
Sums	1 6800	1 2600	1 4700
Loadings	8000	6000	7000

TABLE 4

Forward Doolittle Solution and Beta Weights for Matrix R_{22} with Loadings and Specifics Appended, and Transformation Matrix

Forward Doolittle Solution

	a	b	c	L	S_a	S_b	S_c
A_a	1 0000	4800	5600	-8000	-6000		
R_a	-1 0000	-4800	-5600	8000	6000		
b		1 0000	4200	-6000	0000	-8000	
A_b		7696	1512	-2160	2880	-8000	
R_b		-1 0000	-1965	2807	-3742	1 0395	
c			1 0000	-7000	0000	0000	-7141
A_c			6567	-2096	2794	1572	-7141
R_c			-1 0000	3192	-4255	-2394	1 0874

Beta Weights, B_3

	a	b	c
L	5166	2180	3192
S_a	9778	-2906	-4255
S_b	-3875	1 0865	-2394
S_c	-5064	-2137	1 0874

Multiplier Matrix, M_2

a	666667	0	0
b	0	1 000000	0
c	0	0	8333333

Transformation Matrix, $T_1 = B_3M_2$

L	3444	2180	2660
S_a	6519	-2906	-3546
S_b	-2583	1 0865	-1995
S_c	-3376	-2137	9062

Factor Structure: Second Order; $\left[{}_2\Sigma C_{icj}\right]'\left[T_1\right]' = F_2$

C	L	S_a	S_b	S_c	h^2	$1-h^2$	Specifics
A	6400	4801	0000	0000	6401	3599	5999
B	5600	4201	0001	-0001	4901	5099	7141
C	2400	0001	3200	-0001	1600	8400	9165
D	3600	0000	4800	0000	3600	6400	8000
E	6300	0000	0000	6427	8100	1900	4359
F	2100	0000	0000	2142	0900	9100	5939

TABLE 5

Transformation Matrix and Final Hierarchical Factor Structure for the Problem

Transformation Matrix, $T_2 = B_2{}'M_1$

	A	B	C	D	E	F
L	1930	1559	0588	1544	2692	0202
S_a	2672	2158	-0389	-1023	-1782	-0134
S_b	-0471	-0382	1999	5247	-0656	-0049
S_c	-1305	-1053	-0397	-1046	5672	0425
A	5606	-3388	-0316	-0830	-1446	-0108
B	-3085	8279	-0234	-0611	-1062	-0080
C	-0341	-0277	8240	-2237	-0476	-0035
D	-0586	-0474	-1464	1 0047	-0817	-0062
E	-0865	-0698	-0263	-0693	4131	-0918
F	-0132	-0106	-0040	-0106	-1866	7923

Factor Structure: First Order,

$\left\{{}_1\Sigma C_{icj}\right]' \cdot \left[T_2\right]' = F_1$

	L	S_a	S_b	S_c	A	B	C	D	E	F
1	5120	3840	-0001	0001	4799	-0001	-0001	0000	0001	0000
2	5759	4320	-0001	0001	5398	-0001	-0001	0001	0000	0000
3	3920	2940	-0001	0001	0000	4998	-0001	0000	0001	0000
4	3360	2520	-0001	0001	0000	4284	-0001	0000	0000	0001
5	1919	0001	2560	0001	0000	-0001	7331	0000	0001	0000
6	0959	0001	1280	0000	0000	0000	3666	0000	0000	0000
7	2519	0000	3360	0000	0000	0000	-0001	5601	0000	0000
8	0720	0000	0960	0000	0000	-0001	0000	1600	0000	0000
9	5669	0000	0000	5784	0000	0000	-0001	0000	3923	0001
10	3150	0000	0000	3214	0000	0000	-0001	0001	2179	0000
11	1260	0000	0001	1285	0000	0000	0000	0000	0000	5723
12	1470	0000	0000	1500	0001	-0001	0000	-0001	0000	6677

formation matrix is used to postmultiply the transpose of the second-order cluster sums from Table 2, which yields the factor structure for the second-order domain. The general factor is labeled L and the three sub-general factors are labeled S_a , S_b , and S_c .

Table 5 presents some of these same steps for the transformation into the first-order domain with loadings for the original variables. The first-order transformation matrix (T_1), obtained by a Doolittle solution using first-order cluster intercorrelations, first-order cluster loadings (including specifics) in the second-order domain as criteria, and the diagonal multiplier matrix, is shown. In addition the final loadings on the General, the three sub-generals, and the six group factors are obtained for the twelve original variables by postmultiplying the transpose of the matrix of first-order cluster sums for each test by the transpose of the first-order transformation matrix.

Discussion of Results

Inspection of the final factor loadings in Table 5 discloses that the six group factors possess simple structure with respect to one another and that the three sub-general factors have this same property among themselves. Comparison with the final results obtained by Schmid and Leiman shows no discrepancy greater than plus or minus .0001.

It has been demonstrated that successive applications of the multiple group factor method to the original correlation table and resulting cluster correlation tables at higher orders will, after proper back solution, yield identical results as the Schmid-Leiman approach.

The chief virtue of the present method is that it completely eliminates the rotation problem by using cluster correlations directly rather than (1) rotating to oblique simple structure at each stage and (2) using the inverse of the reference vector matrix to go to the next higher order. It even eliminates the need for the original oblique cluster correlations, and the rotation of these to orthogonality prior to such rotation, since it uses the cluster sums directly.

The hierarchical solution, being completely orthogonal, is easily used to get the residual table. It also presents all factors from all "orders" in a single table so that the whole picture can be seen at a glance. Its group factors are comparable to the Thurstone primary factors except that the present factors are orthogonal. Its sub-general factors bear the same relationship to Thurstone's "second-order" factors, and the general factor, in this instance, to a Thurstone "third-order" factor. Thus the interpretation of the factors would be unchanged—with possibly better interpretation of the higher order factors since the projections of all of the original tests are now explicit on all orders of factors rather than merely on the first-order factors. Since the two types of solution are mathematically equivalent, the question of whether factors are "really" oblique or orthogonal is unanswered.

REFERENCES

[1] Schmid, J. and Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika*, 1957, 22, 53-61.

[2] Thomson, G. H. *The factorial analysis of human ability*. New York: Houghton Mifflin, 1948.

[3] Wherry, R. J. and Winer, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, 18, 161-179.

Manuscript received 1/27/58

Revised manuscript received 6/27/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

A GENERALIZATION OF THE BUSH-MOSTELLER MODEL WITH SOME SIGNIFICANCE TESTS*

MARY I. HANANIA†

UNIVERSITY OF CALIFORNIA, BERKELEY

A generalization of the Bush-Mosteller learning model is proposed in connection with two-alternative learning situations with continuous reinforcement. The problem is to test the hypothesis that reward and non-reward are equally effective in promoting learning. Statistically this reduces to testing a hypothesis about the value of a single parameter θ, while a set of other parameters remains unspecified. The test presented has the property of being asymptotically locally most powerful among all tests of the same size and asymptotically similar. The application of the test is illustrated.

This paper is concerned with testing the effectiveness of reward in simple learning situations of the following type. A subject is observed over a sequence of trials, on each of which he must choose between two mutually exclusive responses, A and B. When he responds by choosing A he receives a reward, and when he chooses B he either receives no reward or is punished. Experiments indicate whatever choice the subject makes on successive trials he eventually learns to make response A. Both a rewarded and an unrewarded response may contribute to the learning process, but the two experiences may differ in their *effectiveness* towards that end. The problem is to determine from a set of experimental data whether reward or non-reward has a greater effect on the subject's learning of that response.

A typical experiment in this class would be running a hungry rat through a simple T-maze. After going through a short runway, the rat has to choose between turning right (response A) or left (response B); the right branch leads to a box containing food (reward), the left to an empty box (non-reward). These same conditions—the hunger drive and the outcomes of the two responses—prevail on all trials.

One can also set up the maze with food on one branch and an electric shock (punishment) on the other. Or, to take a possible variation out of many, one might compare the effectiveness of different degrees of reward by placing a large amount of food in one box and a small amount in the other.

In some of these experiments, there may be doubt as to the exact nature of the particular response to be learned by the subject, whether, for instance, a hungry rat learns to *avoid* one box because of an electric shock in it or to *seek* the other because it contains food. Such distinctions are not made here; rather, responses A and B are defined in terms of the subject's overt behavior, then the effects of the outcomes of A and B are examined. In the case of the hungry rat finding a food box on turning right and an electric shock on turning left, a right turn is classified as response A and a left turn as response B. The question is whether food or shock is more effective in encouraging the rat to make response A on future trials. If a satisfactory answer to this question is obtained, it may then be desirable to redefine responses A and B. But the problem of interpretation is left strictly within the domain of the experimenter.

---

*I would like to express my gratitude to Professors J. Neyman and E. L. Scott, Department of Statistics, University of California, for their constant assistance and encouragement throughout the research that led to this paper and during its preparation, and to Professor F. W. Irwin, Department of Psychology, University of Pennsylvania, for his many helpful suggestions and comments.

†Presently at the American University of Beirut, Beirut, Lebanon.

Learning situations of this nature were among those treated by Bush and Mosteller [1] in the application of their basic model for learning. The model presented and used in this paper was actually suggested by the Bush-Mosteller approach, but allows for more generality in the trial-to-trial transitions.

The Model

Assumptions

The two-alternative learning situation described above consists of a sequence of trials on each of which the subject makes a choice between two responses, A (the response to be learned) and B. Consider the behavior of this subject at the start of trial j in an experiment. While several factors may influence his decision in choosing between A and B, they all combine to give him some degree of preference for one response over the other. This preference may be expressed in terms of the probability with which he chooses one of the two responses. Assume that he chooses B with probability q_j and A with probability $1 - q_j$.

When a decision is reached, the subject acts by making one of the two responses, and his behavior at this stage of the trial can be observed and recorded by the experimenter. For instance, the experimenter may record the result of the subject's choice on the trial j as x_j, where x_j is 1 if the subject makes response A, zero if he makes response B.

With every sequence of trials $1, 2, \cdots, j, \cdots$, describing the behavior of the subject over these trials two other sequences may be connected: the sequence of probabilities $q_1, q_2, \cdots, q_j, \cdots$, of making response B, and the sequence of outcomes $x_1, x_2, \cdots, x_j, \cdots$, where $x_j = 1$ or 0.

Assume that the change in the probability of a subject's making response B from one trial to the next will in general depend on the experience of the subject between the two choices; in other words, assume that, for any j, the probability q_{j+1} depends on x_j as well as on q_j. In particular, given the value of q_j and the outcome of the trial j, the probability of response B on trial $j + 1$ is

$$q_{j+1} = \begin{cases} w_{j+1}q_j \text{ if the subject was not rewarded on trial } j, \\ \theta w_{j+1}q_j \text{ if the subject was rewarded on trial } j, \end{cases}$$

where w_{j+1} is a constant depending upon the combination of factors other than the presence or absence of reward which influence the change from q_j to q_{j+1}, and where θ is a constant such that $0 < \theta \leq 1$. Since the outcome of response A is always reward while the outcome of response B is non-reward, this means that

$$q_{j+1} = \begin{cases} w_{j+1}q_j \text{ if the subject chose } B \text{ on trial } j, \\ \theta w_{j+1}q_j \text{ if the subject chose } A \text{ on trial } j. \end{cases}$$

Using the definition of x_j, this relation can be written more conveniently as

$$(1) \qquad q_{j+1} = \theta^{x_j} w_{j+1} q_j .$$

Unless $\theta = 1$, the probability q_{j+1} depends on whether or not the subject was rewarded on the trial j. In fact, $\theta < 1$ implies that after being rewarded the subject has a smaller chance of making response B than after a non-reward trial. This finding would support the assertion sometimes made that reward has a stronger positive effect on learning. On the other hand, $\theta = 1$ implies that the effects of reward and non-reward on the behavior of the subject are identical.

Iterating relation (1),

$$\begin{aligned} q_2 &= \theta^{x_1} w_2 q_1 , \\ q_3 &= \theta^{x_2} w_3 q_2 = \theta^{x_1+x_2} w_2 w_3 q_1 , \\ q_4 &= \theta^{x_1+x_2+x_3} w_2 w_3 w_4 q_1 , \end{aligned}$$

and in general,

$$q_j = \theta^{x_1+\cdots+x_{j-1}} w_2 \cdots w_j q_1 \qquad (j = 2, 3, \cdots).$$

u_1 may be written for q_1, then u_2 for $w_2 q_1$, etc., since each of the factors is an unknown parameter. In general,

$$u_j = w_2 \cdots w_j q_1 .$$

Now

$$q_j = \theta^{x_1+\cdots+x_{j-1}} u_j \qquad (j = 1, 2, \cdots),$$

where x_0 is defined as zero, and u_j is a constant such that $0 < u_j \leq 1$. The u_j must remain in the unit interval because q_j is a probability; the q_j are taken to be strictly positive to avoid having all probabilities q_{j+k} after the trial j equal to zero no matter what the subject learns.

Since the exponent of θ is the number of rewarded trials among the first $j - 1$, $\theta < 1$ again implies that the occurrence of reward before trial j decreases the probability of response B on that trial. On the other hand, $\theta = 1$ implies that on trial j, q_j has a fixed value u_j, no matter how many previous trials were rewarded. In the remaining part of this section, a set of random variables will be defined with distribution depending on θ. The question of whether or not reward has a greater effect than non-reward on learning will then reduce to the problem of testing the statistical hypothesis that $\theta = 1$ against the alternative that $\theta < 1$.

Suppose that the experiment consists of m consecutive trials for each subject and that it is run independently on r different subjects. Define a set of rm random variables X_{ij} $(i = 1, \cdots, r; j = 1, \cdots, m)$ by

$$X_{ij} = \begin{cases} 1 \text{ if subject } i \text{ makes response } A \text{ on trial } j, \\ 0 \text{ if subject } i \text{ makes response } B \text{ on trial } j. \end{cases}$$

In the same way as above, let

$q_{ij} = P$ [subject i makes response B on trial j | outcome of previous trials], where $P[E_1 \mid E_2]$ denotes the conditional probability of event E_1, given E_2. Then

$$q_{ij} = P[X_{ij} = 0 \mid x_{i1}, \cdots, x_{i,j-1}].$$

Define a set Ω of admissible hypotheses on the distribution of the random variables with the following assumptions.

(i) All subjects have the same initial positive probability on the first trial, that is, $q_{i1} = u_1 > 0$ $(i = 1, \cdots, r)$.

(ii) Subjects act and learn independently of each other, that is, the X_{ij} are completely independent for different i.

(iii) Assume

$$q_{ij} = \theta^{\sum_{k=1}^{j-1} x_{ik}} u_j \qquad (i = 1, \cdots, r; j = 1, \cdots, m), \tag{2}$$

where $0 < \theta \leq 1$, $0 < u_j \leq 1$, and $x_{i0} = 0$.

The third assumption states that for subject i the probability that $X_{ij} = 0$ depends on the sum $\sum_{k=1}^{j-i} x_{ik}$, which is the total number of A responses made by that subject before trial j. It also depends on θ and, for each j, on a different parameter u_j. However, the u_j are the same for all subjects.

Relation to the Bush-Mosteller Model

A direct application to this problem of the basic model introduced by Bush and Mosteller ([1], ch. 1) results in an expression for q_{ij} identical with (2) in form but involving only three instead of $m + 1$ unknown parameters. Bush and Mosteller assume that, for the subject i and any pair of probabilities

q_{ij} and $q_{i,j+1}$, the transition in probability from q_{ij} to $q_{i,j+1}$ can be expressed in terms of the outcome of the trial j and some constant parameters of the learning situation that do *not* depend on j. In particular, their "fixed-point form" ([1], p. 29) assumes that

$$q_{i,j+1} = \alpha q_{i,j} + (1 - \alpha)\lambda,$$

where α and λ are constant parameters of the event occurring on the trial j. In this learning situation, where response A is always rewarded and response B never rewarded, the "fixed-point" parameter λ([1], p. 60) is zero, and the basic assumption reduces to the expression

$$q_{i,j+1} = \begin{cases} bq_{ij} \text{ if subject } i \text{ chose } B \text{ on trial } j, \\ \theta bq_{ij} \text{ if subject } i \text{ chose } A \text{ on trial } j, \end{cases}$$

for any $j = 2, \cdots, m$, where $0 < \theta \leq 1$, $0 < b < 1$, so that θb is their reward and b their non-reward parameter.

Using the random variable X_{ij} already introduced, this probability for the subject i is

$$q_{i,j+1} = \theta^{x_{ij}} bq_{ij} \qquad (j = 2, \cdots, m).$$

This relation is identical with (1) in form, with the constant b replacing w_{j+1} for all j. Thus, it is a special case of (1) where the trial-to-trial transitions in the probability q_j are more restricted than in our general model.

This iterative relation can again be used to deduce a general expression for q_{ij} as in the above section. In this case,

$$q_{i2} = \theta^{x_{i1}} bq_{i1},$$

$$q_{i3} = \theta^{x_{i2}} bq_{i2} = \theta^{x_{i1}+x_{i2}} b^2 q_{i1},$$

and, in general,

$$q_{ij} = \theta^{x_{i1}+\cdots+x_{i,j-1}} b^{j-1} q_{i1} .$$

Assuming that $q_{i1} = q > 0$ $(i = 1, \cdots, r)$, which is equivalent to assumption (i) in the model,

$$q_{ij} = \theta^{\sum_{k=1}^{j-1} x_{ik}} b^{j-1} q. \tag{3}$$

Equation (3) is the same as (2), with the additional restriction that each $u_j = b^{j-1}q$. The parameters $u_1, \cdots, u_m$ are therefore replaced in this special case by a decreasing geometric sequence $q, bq, \cdots, b^{m-1}q$. It is possible to impose a restriction on the parameters u_j in the model by assuming that $u_1 > u_2 > \cdots > u_m$, which would bring it closer to the Bush-Mosteller version. It can be shown ([2], p. 19) that the test proposed in the next section would still hold in this partially restricted case. However, the results given below were obtained without an ordering restriction being imposed upon the parameters.

Testing the Hypothesis that $\theta = 1$

Theoretical Considerations

In order to state the testing problem more precisely, the distribution of the random variables under the set Ω of admissible hypotheses is required. The general form for this distribution can be obtained from (2), which gives the conditional probability $P[X_{ij} = 0 \mid x_{i1}, \cdots, x_{i,j-1}]$ for any one of the random variables X_{ij} as

$$P[X_{ij=0} \mid x_{i1}, \cdots, x_{i,j-1}] = \theta^{\sum_{k=1}^{j-1} x_{ik}} u_j \quad (i = 1, \cdots, r; \quad j = 1, \cdots, m).$$

Starting with this relation, one can verify that the joint probability distribution of the rm random variables is the product

$$\begin{aligned} P[X_{ij} &= x_{ij}; \quad \text{all } i \text{ and } j \mid \Omega] \\ &= \prod_{i=1}^{r} \prod_{j=1}^{m} (1 - \theta^{\sum_{k=1}^{j-1} x_{ik}} u_j)^{x_{ij}} (\theta^{\sum_{k=1}^{j-1} x_{ik}} u_j)^{1-x_{ij}}, \end{aligned} \tag{4}$$

where $0 < \theta \leq 1$ and $0 < u_j \leq 1$ $(j = 1, \cdots, m)$.

The hypothesis to be tested is the assertion that reward and non-reward have identical effects on the behavior of the subject. In terms of the parameters of the distribution, this is the hypothesis,

$$H: \theta = 1,$$

so that in this case the joint probability distribution (4) reduces to

$$P[X_{ij} = x_{ij}; \quad \text{all } i \text{ and } j \mid H] = \prod_{j=1}^{m} (1 - u_j)^{\sum_{i=1}^{r} x_{ij}} u_j^{r - \sum_{i=1}^{r} x_{ij}}, \tag{5}$$

where $0 < u_j \leq 1$ $(j = 1, \cdots, m)$.

Notice that the hypothesis tested, H, is a composite hypothesis, since it does not specify the values of the parameters $u_1, \cdots, u_m$. The problem is to find a test of this composite hypothesis against the set of alternatives that $0 < \theta < 1$ at a fixed level of significance α.

Since the distribution of the random variables X_{ij} which is given in (4) and (5) is discrete, it will in general be necessary to use a randomized test in order to make the probability of rejecting the hypothesis when true equal exactly α. With a randomized test, the points of the sample space under consideration are divided into three mutually exclusive sets: ω_1, ω_2, and ω_3. Each point represents a possible matrix of observed values x_{ij} and will be referred to briefly as a sample point x. The hypothesis tested is rejected whenever the observed sample point x falls within ω_1 and is accepted whenever x falls in ω_2. When x falls in ω_3, the hypothesis is either rejected or accepted, the decision to reject or accept in such a case being determined, with probability specified by the test, by some random process completely independent of the experiment producing the observation x, such as the use of a table of random numbers. In order to be more precise, the Lehmann-Stein notation [3] is used. Define any test T in terms of its critical function $\psi(x)$, with $0 \leq \psi(x) \leq 1$, where $\psi(x)$ is the probability with which test T rejects the hypothesis H when the point x is observed. Thus ω_1 is the set where $\psi(x)$ is 1, while ω_2 is the set where $\psi(x) = 0$, and ω_3 is the set where $0 < \psi(x) < 1$. The value of $\psi(x)$ is determined so that $E[\psi(X)|\, H] = \alpha$. Here, α is the probability with which the test T rejects the hypothesis H; it is also known as the *size* of T.

In addition to the size of the test, it is also important to consider its power. In particular, following the theory of hypothesis testing developed by Neyman and Pearson [4], it is desirable that for the composite hypothesis H there is a test T with the following two properties.

(i) It is a *similar* test of size α. In other words, the probability of rejecting the hypothesis H when $\theta = 1$ is equal to α whatever be the values of $u_1, \cdots, u_m$ in Ω. This property can be stated symbolically as

$$E[\psi(X)|\, H] \equiv \alpha \text{ for all possible values of } u_i\,.$$

(ii) The similar test is *most powerful* in the sense that, when $\theta < 1$, the probability of rejecting the hypothesis that $\theta = 1$ be the largest possible among all similar tests of the same size. If H were tested against a *simple* alternative

$$K\colon \theta = \theta', \qquad u_i = u_i',$$

where $0 < \theta' < 1$ and with u_i' a specified set of values of u_i, this property could be stated symbolically as

$$E[\psi(X) \mid K] \geq E[\psi_1(X) \mid K],$$

where $\psi_1(X)$ is the critical function of any other test T_1 having property (i).

Now, if in addition to being similar, T possesses property (ii) independently of the specified values $\theta', u_1', \cdots, u_m'$, then T would be the most powerful similar test against *all* alternatives in Ω. Such a test is known as a *uniformly most powerful* similar test. A uniformly most powerful similar test would be desirable in this problem since the values of θ and $u_1, \cdots, u_m$ are not specified in the alternative against which the test is made. If, on the other hand, the similar test of H which is most powerful against a simple alternative in Ω depends on the value of any of the parameters specified by the alternative, then the test is not most powerful against all the alternatives in Ω, and there exists no uniformly most powerful similar test of H.

In the event that no uniformly most powerful similar test exists, one might be satisfied with a similar test that is *locally best*, in the sense that it will be most powerful in detecting small deviations from unity in θ. Thus, while still requiring that the test define a rejection rule based on the observation x which is independent of the values of $u_1, \cdots, u_m$, one might consider only "local" alternatives with θ in the neighborhood of the value tested, $\theta = 1$.

Unfortunately, however, it has been found that neither a similar test that is uniformly most powerful against all alternatives in Ω nor one that is locally best exists ([2], secs. 3.1, 3.2), as such tests appear to depend on the values of the unknown parameters.

An alternative approach is possible by means of a method for constructing asymptotic tests for composite hypotheses first given by Neyman [5] in the case of a single unknown parameter. The method utilizes an estimate of the unknown parameter and, allowing the sample size r to increase indefinitely, defines a sequence of critical functions $\psi_r(X)$ such that, as $r \to \infty$, the expected value of $\psi_r(X)$ under the hypothesis tested tends to the preassigned level α, independently of the unspecified parameter. A test defined by such a sequence of critical functions is called *asymptotically similar*.

By an extension of Neyman's method to the case where the distribution of the random variables under the hypothesis tested involves more than one unknown parameter, it is possible to construct an asymptotically similar test of the composite hypothesis H: $\theta = 1$. Such a test is given in the next section and will be referred to as test T_1.

An Asymptotically Locally Best Similar Test of H

Following Neyman, the sequence of critical functions $\psi_r(x)$ will be derived from a set of r independently distributed functions of the observations, say $\zeta_1, \cdots, \zeta_r$, through a number of simple transformations and with the use of a set of consistent estimates for the unspecified parameters $u_1, \cdots, u_m$.

The functions $\zeta_1, \cdots, \zeta_r$ must satisfy certain general conditions of finiteness and differentiability ([2], ch. 3.3), which are easily met. However, of all the possible sets of functions $\zeta_1, \cdots, \zeta_r$ satisfying those conditions, one selects a suitable set such that the power function of the resulting asymptotically similar test would have some optimum local property. For this purpose, first consider a modified hypothesis in Ω; in particular, the hypothesis

$$H_0: \theta = 1; \qquad u_1^0, \cdots, u_m^0,$$

where u_j^0 is a specified value of u_j $(j = 1, \cdots, m)$. Now look for the locally best size α test, say T_0, of the hypothesis H_0 against the alternatives

$$K: \theta < 1; \qquad u_1^0, \cdots, u_m^0 .$$

That is, in the class of all size α tests of H_0 against K, the power function of the test T_0 must have numerically the largest slope at $\theta = 1$.

Such a test can be easily obtained from the probability distribution (4) by application of the Neyman-Pearson Fundamental Lemma [4], and is given by

$$\psi(x) = 1 \quad \text{whenever} \quad L > k,$$
$$\psi(x) = 0 \quad \text{whenever} \quad L < k,$$
$$0 \leq \psi(x) \leq 1 \quad \text{whenever} \quad L = k,$$

where

$$L = \sum_{i=1}^{r} \sum_{j=2}^{m} \frac{x_{ij} - (1 - u_j^0)}{1 - u_j^0} \sum_{k=1}^{j-1} x_{ik} , \tag{6}$$

and k is so chosen that $E[\psi(X) | H_0] = \alpha$.

In close relation to L, the set of r functions ζ_i now is defined as

$$\zeta_i = \sum_{j=2}^{m} \frac{x_{ij} - v_j}{v_j} \sum_{k=1}^{j-1} x_{ik} \qquad (i = 1, \cdots, r), \tag{7}$$

where $v_j = 1 - u_j$ for all j. For each i, the function ζ_i is seen to depend on the unknown parameters $u_2, \cdots, u_m$ as well as on the set of observations $x_{i1}, \cdots, x_{im}$, which may be considered as a single observation on the random vector $X_i = (X_{i1}, \cdots, X_{im})$. Under the hypothesis tested $H: \theta = 1$, the joint distribution of the r random vectors $X_1, \cdots, X_r$ is given by (5), from which it is clear that the vectors are independently and identically distributed with the probability distribution

$$P[X_i = x_i \mid u_1, \cdots, u_m ; H] = \prod_{j=1}^{m} v_j^{x_{ij}} u_j^{1-x_{ij}} \qquad (i = 1, \cdots, r). \tag{8}$$

It follows that the functions $\zeta_1, \cdots, \zeta_r$ as defined in (7) must also be mutually independent.

A critical function $\psi_r(x)$ now is derived from the r functions $\zeta_1, \cdots, \zeta_r$ through the following transformations.

(i) For each i, let the standardized value of ζ_i be

$$Z_i = \frac{\zeta_i - E(\zeta_i)}{\sigma(\zeta_i)}.$$

As can be verified, both the expected value $E(\zeta_i)$ and the variance $\sigma^2(\zeta_i)$ of the function ζ_i are finite, and the resulting standardized function is

$$Z_i = \frac{\sum_{j=2}^{m} \left(\frac{x_{ij} - v_j}{v_j} \sum_{k=1}^{j-1} x_{ik} \right)}{\sqrt{\sum_{j=2}^{m} \frac{u_j}{v_j} \left[\left(\sum_{k=1}^{j-1} v_k \right)^2 + \sum_{k=1}^{j-1} v_k u_k \right]}}. \tag{9}$$

(ii) Now introduce the m functions

$$\varphi_j = \frac{\partial}{\partial u_j} \log \; P[X_i = x_i \mid u_1 , \cdots , u_m ; H]$$

for $j = 1, \cdots , m$, and let

$$\phi_j = \frac{\varphi_j - E(\varphi_j)}{\sigma(\varphi_j)}$$

be the standardized value of φ_j . Then each ϕ_j $(j = 1, \cdots , m)$, is a function of X_i and $u_1 , \cdots , u_m$.

(iii) Also define m other functions $R_1 , \cdots , R_m$ by

$$R_j = E(Z_i \phi_j).$$

(iv) Finally, for each i, combine $Z_i , \phi_1 , \cdots , \phi_m$, and $R_1 , \cdots , R_m$ into the function

$$f_i(X_i ; u_1 , \cdots , u_m) = \frac{Z_i - \sum_{j=1}^{m} R_j \phi_j}{\sqrt{1 - \sum_{j=1}^{m} R_j^2}} , \tag{10}$$

and take the average

$$Y_r = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} f_i(X_i ; u_1 , \cdots , u_m).$$

In terms of the original observations,

$$Y_r = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} \frac{\sum_{j=2}^{m} \frac{x_{ij} - v_j}{v_j} \sum_{k=1}^{j-1} (x_{ik} - v_k)}{\sqrt{\sum_{j=2}^{m} \frac{u_j}{v_j} \sum_{k=1}^{j-1} v_k u_k}} . \tag{11}$$

Equation (11) can be derived from (9) and (10) and the definitions of the functions introduced above.

Y_r is clearly a function of the unknown parameters, as well as of the rm observations x_{ij} $(i = 1, \cdots , r; j = 1, \cdots , m)$. Replacing each u_j in Y_r by an estimate u_j^* for $j = 1, \cdots , m$, gives the corresponding function

$$Y_r^* = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} f_i(X_i ; u_1^* , \cdots , u_m^*),$$

whose definition is independent of $u_1 , \cdots , u_m$. A suitable set of estimates can be found by the maximum likelihood method, which gives

$$\frac{c_j}{r} = \frac{1}{r} \sum_{i=1}^{r} x_{ij}$$

as the maximum likelihood estimate of $1 - u_j$.

Therefore take

$$1 - u_j^* = v_j^* = \frac{c_j}{r} \quad \text{whenever} \quad c_j \neq 0,$$

$$v_j^* = \frac{1}{r^2} \quad \text{whenever} \quad c_j = 0.$$

The provision in the latter case is introduced in order to avoid division by zero in Y_r^*.

Substituting these estimates for $u_1, \cdots, u_m$ (or $v_1, \cdots, v_m$) in (11) and simplifying,

$$Y_r^* = \frac{\sum_{i=1}^{r} \sum_{j=2}^{m} \frac{x_{ij}}{c_j} \sum_{k=1}^{j-1} x_{ik} - \frac{1}{r} \sum_{j=1}^{m-1} (m-j)c_j}{\sqrt{\frac{1}{r^3} \sum_{j=2}^{m} \frac{r - c_j}{c_j} \sum_{k=1}^{j-1} c_k(r - c_k)}}, \tag{12}$$

where $c_j = 0$ is replaced by $1/r$.

Under the hypothesis H the probability distribution of this function Y_r^* can be shown to approach the normal with zero mean and unit variance as $r \to \infty$ ([2], ch. 3.3).

Now the test T_1 is defined as follows. Compute Y_r^* from the data; choose a level of significance α and reject the hypothesis H with probability $\psi_r(x)$, where

$$\psi_r(x) = 1 \quad \text{if} \quad Y_r > k_\alpha,$$
$$\psi_r(x) = 0 \quad \text{if} \quad Y_r < k_\alpha,$$
$$0 \le \psi_r(x) \le 1 \quad \text{if} \quad Y_r = k_\alpha,$$

and k_α is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_\alpha} e^{-t^2/2}\, dt = 1 - \alpha.$$

This test has the following properties.

(i) By definition of Y_r^* in (12), $E[\psi_r(X) \mid H]$ is independent of the unspecified parameters $u_1, \cdots, u_m$.

(ii) Since the distribution of Y_r^* approaches $N(0, 1)$ under H as $r \to \infty$, the size of the test $E[\psi_r(X) \mid H] \to \alpha$, the required significance level.

(iii) By comparison of the asymptotic distributions of L in (6) and Y_r^* in (12) the test T_1 may be shown to be asymptotically equivalent to the locally best test T_0.

Therefore test T_1 is asymptotically a locally best test of H against the alternatives K: $\theta < 1$; $u_1, \cdots, u_m$, and asymptotically similar of size α.

Application of Test T_1

Given is a set of rm observations x_{ij} $(i = 1, \cdots, r; j = 1, \cdots, m)$, where r is the number of subjects in an experiment of m trials, with

$$x_{ij} = \begin{cases} 1 \text{ if the subject } i \text{ is rewarded on the trial } j, \\ 0 \text{ if the subject } i \text{ is not rewarded on the trial } j. \end{cases}$$

To test the hypothesis H: $\theta = 1$ (reward and non-reward are equally effective) against the alternatives that $\theta < 1$ (reward is more effective) at the level of significance α, apply the following rule.

(i) From the data, form the sums

$$c_j = \sum_{i=1}^{r} x_{ij} \qquad (j = 1, \cdots, r).$$

Then c_j is the total number of subjects rewarded on the trial j. Let

$$c'_j = c_j \quad \text{whenever} \quad c_j \neq 0,$$
$$c'_j = \frac{1}{r} \quad \text{whenever} \quad c_j = 0.$$

(ii) Compute

$$M = \sum_{i=1}^{r} \sum_{j=2}^{m} \frac{x_{ij}}{c'_j} \sum_{k=1}^{j-1} x_{ik} ,$$

$$C' = \frac{1}{r} \sum_{j=1}^{m-1} (m - j) c'_j ,$$

$$C'' = \sqrt{\frac{1}{r^3} \sum_{j=2}^{m} \frac{r - c'_j}{c'_j} \sum_{k=1}^{j-1} c'_k (r - c'_k)} ,$$

and combine these three quantities to obtain

$$Y = \frac{M - C'}{C''}. \tag{13}$$

(iii) From a table of the normal distribution function find the upper α point, k_α, defined by

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_\alpha} e^{-t^2/2}\, dt = 1 - \alpha. \tag{14}$$

Then reject H with probability $\psi(x)$, where

$$\begin{aligned} \psi(x) &= 1 \quad \text{if} \quad Y > k_\alpha , \\ \psi(x) &= 0 \quad \text{if} \quad Y < k_\alpha , \\ 0 \leq \psi(x) &\leq 1 \quad \text{if} \quad Y = k_\alpha . \end{aligned} \tag{15}$$

The probability of rejecting the hypothesis H when Y as computed from the data is exactly equal to k_α should be chosen so as to make $E[\psi(X) \mid H] = \alpha$.

Illustration

The use of the test T_1 is illustrated on data from an experiment on prediction by F. W. Irwin and L. Dyckman*, in which the subjects were 20 women undergraduates. At the start of each trial the subjects were asked to predict whether or not a light would go on. The light then went on, thus it may be said that a subject who had predicted it would go on was "rewarded," or agreed with. Since the light went on every time regardless of the subject's choice, this is a case of continuous reinforcement.

Table 1 shows the outcome of the experiment in the form of a matrix of observations for 20 subjects and 22 trials, with an entry of one for a correct prediction and zero for an incorrect prediction. In this experiment, $r = 20$ and $m = 22$. For each j ($j = 1, \cdots, 22$),

TABLE 1

Data of F.W. Irwin and L. Dyckman from a "Humphreys-type" Experiment with Continuous Reinforcement *

Subjects (i)	Trials (j)																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	0	1	0	0	1	1	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1
3	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
4	1	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
10	1	0	1	0	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
11	1	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
12	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1
14	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
17	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
19	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Total (c_j)	18	10	9	9	8	14	13	14	18	17	18	19	19	17	19	20	19	20	20	19	19	20

* An entry of 1 indicates $x_{ij}=1$ (reward), a zero entry indicates $x_{ij}=0$.

$$c_j = \sum_{i=1}^{20} x_{ij} ,$$

from which are obtained

$$c_j' , \quad \frac{1}{c_j'} , \quad 20 - c_j' , \quad \frac{20 - c_j'}{c_j'} ,$$

---

*Personal communication from F. W. Irwin.

and

$$\sum_{k=1}^{j-1} c_k'(20 - c_k') \qquad (j = 2, \cdots, 22).$$

Then compute

$$M = \sum_{i=1}^{20} \sum_{j=2}^{22} \frac{x_{ij}}{c_j'} \sum_{k=1}^{j-1} x_{ik} ,$$

$$C' = \frac{1}{20} \sum_{j=1}^{21} (22 - j)c_j' ,$$

$$C'' = \left[\frac{1}{8000} \sum_{j=2}^{22} \frac{20 - c_j'}{c_j'} \sum_{k=1}^{j-1} c_k'(20 - c_k')\right]^{1/2},$$

as shown in Table 2. Then, by (13),

$$Y = \frac{M - C'}{C''} = 4.186.$$

TABLE 2

Table Illustrating the Computing of Y in the Irwin-Dyckman Data

j	$22-j$	c_j'	$20-c_j'$	$\frac{20-c_j'}{c_j'}$	$\sum_{k=1}^{j-1} c_k'(20-c_k')$	$\sum_{i=1}^{20} x_{ij} \sum_{k=1}^{j-1} x_{ik}$	$\frac{1}{c_j'} \sum_{i=1}^{20} x_{ij} \sum_{k=1}^{j-1} x_{ik}$
1	21	18	2	.1111			
2	20	10	10	1.0000	36	9	.9000
3	19	9	11	1.2222	136	14	1.5556
4	18	9	11	1.2222	235	17	1.8889
5	17	8	12	1.5000	331	19	2.3750
6	16	14	6	.4286	415	37	2.6429
7	15	13	7	.5385	506	47	3.6154
8	14	14	6	.4286	590	64	4.5714
9	13	18	2	.1111	626	87	4.8333
10	12	17	3	.1765	677	100	5.8824
11	11	18	2	.1111	713	116	6.4444
12	10	19	1	.0526	732	143	7.5263
13	9	19	1	.0526	751	162	8.5263
14	8	17	3	.1765	802	160	9.4118
15	7	19	1	.0526	821	198	10.4211
16	6	20	0	0	821	222	11.1000
17	5	19	1	.0526	840	236	12.4211
18	4	20	0	0	840	261	13.0500
19	3	20	0	0	840	281	14.0500
20	2	19	1	.0526	859	289	15.2105
21	1	19	1	.0526	878	302	15.8947
22	0	20	0	0	878	339	16.9500

$M = 169.2711$

$C' = \frac{3340}{20} = 167.0$ $\quad C'' = \left(\frac{2355.7072}{8000}\right)^{1/2} = 0.5426$ $\quad Y = \frac{2.2711}{0.5426} = 4.1856$

To test the hypothesis that $\theta = 1$ (reward and non-reward, or agreement and contradiction, are equally effective) at the level of significance .05, the rule as given by (15) is to reject the hypothesis tested if $Y \geq k_{.05} = 1.645$. Within the framework of the model, the result of the computations in Table 2 would therefore lead to rejection of the hypothesis that $\theta = 1$, which implies acceptance of the alternative that reward (making a correct prediction) had a stronger effect on learning in this experiment.

Since the above example is given here primarily for purposes of illustrating the computations, it is not necessary to proceed further into the interpretations of the statistical test and resulting conclusions. It need hardly be emphasized that the validity of the statistical model must be carefully considered in each individual experiment before attempting to draw conclusions from the application of the test.

Testing Against Other Alternatives

Up to now, the question under consideration has been whether the effect of reward is equal to or greater than that of non-reward. Thus the set Ω of admissible hypotheses allows θ to assume values either equal to or less than unity, but never greater. However, in some experiments there is a question whether non-reward, or perhaps punishment, has a stronger effect than reward. To allow for this possibility, assume a larger set of admissible hypotheses, say

$$\Omega_1 : 0 < \theta < \theta_1$$

where $\theta_1 > 1$ is defined so that the probability

$$q_{ij} = \theta^{\sum_{k=1}^{j-1} x_{ik}} u_j$$

remains less than or equal to one for all j. This is achieved by letting $\theta_1 = 1/\max u_j$. In this new set Ω_1 , the results of a preceding section may be immediately applied to the case of testing the hypothesis

$$H_2 : \theta = 1$$

against the alternatives $1 < \theta < \theta_1$. Notice that this situation is a mirror of the situation originally discussed above, namely the testing of $H: \theta = 1$ against alternatives in which $0 < \theta < 1$. It will be natural to expect, therefore, that the test of H_2 be the exact mirror of the test T_1 of H, given in (15). In fact, reversing the inequalities of (15), a test of H_2 against alternatives with $\theta > 1$ is obtained. By considerations of symmetry this test of H_2 can be shown to have the same optimum asymptotic properties as does the test T_1 of H against alternatives with $\theta < 1$. This test of H_2 , to be referred to as T_2 , is given in terms of its critical function $\psi(x)$ as

$$\begin{aligned} \psi(x) &= 1 \quad \text{if} \quad Y < k_\alpha , \\ \psi(x) &= 0 \quad \text{if} \quad Y > k_\alpha , \\ 0 \le \psi(x) &\le 1 \quad \text{if} \quad Y = k_\alpha , \end{aligned} \tag{16}$$

where Y and k_α are defined as in (13) and (14), respectively.

A Note on the Use of the Asymptotic Tests in Small Samples

By definition, the asymptotic test has the property that its size, or the probability of its Type I error (rejecting the hypothesis tested when true), approaches the preassigned value α as $r \to \infty$. In practice, therefore, one has at best only an approximate test whose size gets closer and closer to α as the number r of subjects in the experiment increases. The question then arises as to how large must the number of subjects be in order that the probability of Type I error of the test be within some reasonably small interval around α.

A partial answer to this question can be found in a study of the behavior of the asymptotic test T_1 in some selected cases in which the sample size r is small ([2], sec 5). To some extent the approximation seems to depend on the sizes of the statistics c_j in the data, where $c_j = \sum_{i=1}^{r} x_{ij}$ $(j = 1, \cdots, m)$, so that $0 \leq c_j \leq r$. For the case $r = 10$, $m = 4$, the asymptotic test gave fairly good approximations when none of the c_j $(j = 1, \cdots, 4)$ had an extreme value like 0, 9 or 10. For $\alpha = .05$, the actual probability for Type I error in such cases appeared to lie between .01 and .08, being very close to .05 in most examples. The approximations became progressively worse when one or more c_j $(j = 1, \cdots, 4)$ were given values near zero or ten.

As expected, the asymptotic test gives a better approximation to the desired size of the test as the number of subjects r increases. There are also indications of improvements in the approximation as a result of an increase in the length m of the experiment.

REFERENCES

[1] Bush, R. R. and Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.
[2] Hanania, M. Some statistical tests of hypotheses on the effect of reward in learning. Unpublished doctoral dissertation, Univ. California, Berkeley, 1957.
[3] Lehmann, E. L. and Stein, C. Most powerful tests of composite hypotheses: I. Normal distributions. *Ann. math. Statist.*, 1948, **19**, 495-516.
[4] Neyman, J. and Pearson, E. S. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans., Ser. A*, 1933, **231**, 289-337.
[5] Neyman, J. Sur une famille de tests asymptotiques des hypothèses statistiques composées. *Trabajos de Estadística*, 1954, **5**, 161-168.

Manuscript received 4/23/58
Revised manuscript received 7/9/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

SOME IMPLICATIONS OF THE LOGICAL CALCULUS FOR EMPIRICAL CLASSES FOR SOCIAL SCIENCE METHODOLOGY*

WILLIAM C. SCHUTZ†

HARVARD UNIVERSITY

An exposition of a calculus for empirical classes (CEC), one of the few attempts by logicians to deal with the problem of constructs and indicators, is presented. The CEC provides the groundwork for a formal structure for the situation in which individuals have a degree of membership in various classes rather than having either membership or nonmembership—a situation nearly always true in empirical research. The CEC is presented and its relation to various social science concepts is mentioned. An application of the CEC model to latent structure analysis (LSA) suggests alternatives to the local independence assumption including one called the local scale assumption, which has a close relation to a Guttman scale.

Following is an attempt to interpret for social scientists an article by Kaplan and Schott [3] on a calculus for empirical classes, and to suggest some relations of this article to social science research. The article was written by two philosophers. It was intended as a follow-up article to an earlier article by Kaplan [2].

The Kaplan and Schott work is a generalization of what is known to logicians as the calculus of classes or sets. A class may be defined as the totality of objects having a certain property. An object is spoken of as belonging to, or being a member of a given class as "Lazarus belongs to the Socialist Party" or "Mendel is an Elk." Perhaps the reader, especially if he is a social scientist, has felt the inadequacy of the conception of class membership to his own problems, an inadequacy which rests chiefly in the fact that these logical operations require, for their applicability, a certitude as to the class membership or nonmembership of empirical cases, a condition seldom achieved in actual research work. Only in trivial cases can it be said with certainty that "A is a schizophrenic" or "B is upward mobile."

Kaplan and Schott (hereafter referred to as K-S) set out to construct a logical system which is precisely analogous to the conventional calculus (which they call the calculus of sets reserving the term calculus of classes for the new system which they set out to develop) but differs by introducing a notion of degree or weight, instead of the certainty, that an individual either does or does not belong to a given class (set). The combination of qualities from which his class is to be inferred they call the individual's profile. (The terms profile and weight will be given precise definitions in a following section.)

---

*This work was done while the author was an S.S.R.C. Fellow at Columbia University working on a project in social science methodology. Much is owed to Professor Paul Lazarsfeld for frequent discussions and suggestions both prior to and during the writing of this paper. Herbert Menzel also contributed valuable critical comments. Professor Abraham Kaplan also aided by reading the manuscript and offering helpful comments.

†Now at the University of California, Berkeley.

The calculus of empirical classes (hereafter referred to as CEC) is felt to be worth exploring for social scientists for the elucidation of what are called constructs, intervening variables, latent classes, etc. The authors attempt to develop a logical calculus (formal model) of classes that approximate the empirical classes commonly dealt with in research. As such it is one of the few attempts by philosophers to deal with this area. The calculus of probability is an integral part of the Kaplan-Schott development, just as it is more and more proving to be in empirical work.

The CEC also has implications for the concept of operational definition. The probabilistic character of the relation between a concept and the various measures of it is explicated by the CEC. Perhaps the term *indicator definition* will prove more useful than operational definition. This point is discussed at greater length in the original article by Kaplan [2].

> Our problem can be phrased: to construct a calculus (an uninterpreted logical system) which will provide a more adequate explication of classes in their scientific use than is afforded by the conventional "calculus of classes" which we prefer to call the calculus of sets.
>
> The procedure of the paper is to construct . . . entities which have a degree of vagueness characteristic of actual (empirical) classes, but which, when this degree is minimal, correspond to precise sets (those classes for which every element either is or is not a member).

It should be noted that the attempt is to construct a calculus, i.e., a logical system not interpreted in terms of empirical correlates, although the authors have in mind a specific interpretation, namely, empirical classes. In this exposition, this interpretation will frequently be made so that the formal structure will not obscure the relevance of the concepts being offered to social science.

Preliminary Ideas and Terminology

K-S defined the *field* (of inquiry) as the set of individuals with which a particular inquiry is concerned. In the social sciences the individuals are primarily individual persons or groups.

Articulation is the partitioning or dividing of a field into a set of *categories*, each of which exhaustively divides a set of individuals into mutually exclusive sets according to their specific *qualities* in the categories in question. An articulation then is a set of sets (the categories) of exclusive sets (the qualities). For example, a population of army officer candidates (the field) may be articulated by use of the variables or attributes (categories), ascendance, intelligence, age, and length of service. Within each category there are certain values which can be assigned to each officer (thus dividing them into mutually exclusive sets), e.g., high ascendance, low ascendance; under 20 years of age, 21 to 25, 26 to 30, over 31 years of age. A similarity should be noted between the notion of an articulated field and the notion of property space as elaborated by Lazarsfeld and Barton [4].

A category must be exhaustive, i. e., it must assign one quality to each individual, and since the qualities are mutually exclusive, the category assigns only one quality to each individual. Categories are what are frequently called *variables*, *attributes*, or *dimensions* in social science. They would include for people: age, sex, amount of education, dominance, dependence, friendliness, socio-economic status, etc., and for groups: cohesion, democratic, productive, satisfaction, etc.

Some categories may have few qualities and others many, where qualities are mutually exclusive sets of particular categories. Qualities in the social sciences are often called *values* of a variable, *scale points*, *discriminable classes* along a dimension, etc. Examples are: high, low; very much, some, very little; higher, same as, lower; agrees, disagrees. In every case they are mutually exclusive, i.e., no individual can have more than one quality. Refinements of methods of observation can be taken into account in terms of an increase in the number of qualities discriminated.

To clarify the next terms introduced, an example to be used throughout this presentation will be constructed. The *field* is considered to be all army officer candidates at a given army post numbering 1,000, the size of the field. This field will be articulated in two ways, called for simplicity psychological and sociological (see Table 1). (In passing, it may be noted how two different articulations of the same field may be constructed, that each category applies to every individual, and that the qualities within each category are mutually exclusive.)

A *profile* is a set of qualities, one from each category in an articulation. A profile describes an individual completely with respect to a context of inquiry. It corresponds to what in social science is often called a complete descriptive category, or typology, or a point in a property space. In sample articulations the following would be examples of profiles, or complete descriptions of individuals: Independent-Friendly-Low Participator (IFL); 21 to 25-High School-Working Class-Father Army Man (25HWA). It should be noted that the profile is the name for the *descriptive category*—the set of qualities; it is not the name for the individuals who fall into the category. The set of individuals characterized by a profile is called the *unit cell*. The unit cell represents the limits of discrimination possible within an articulation. Individual members of a unit cell are indistinguishable within the articulation. They must be identified as distinct individuals by criteria external to the articulation.

In the examples, all those officer candidates who are, e.g., Dominant-

TABLE 1

Two Articulations of the Field "Officer"

Categories	Qualities	Abbreviations
	Psychological Articulation	
Dominance	Dominant - Independent - Submissive	D, I, S
Friendliness	Too friendly - Friendly - Superficially friendly - Not friendly	T, F, S, N
Participation	High participation - Low participation	H, L
	Sociological Articulation	
Age	18-20, 21-25, 26-30, 31-35, 36-40	20,25,30,35,40
Education	Grade School - High School - College	G, H, C
Social Class	Lower - Working - Middle - Upper	L, W, M, U
Father's Military Status	Regular Army Man - Not Army Man	A, N

Too Friendly-High Participators (DTH) constitute a unit cell. Since there are no other categories in the articulation to use for discriminating the members of the unit cell, they are indistinguishable. Criteria external to the articulation must be used to distinguish them, e.g. proper names.

To consider a group of individuals alike in some ways but not in others a *sub-profile* is used. This is a profile of a sub-set of categories of the articulation, in other words, a set of qualities one from each of some categories but not all of an articulation. The classification 20-College-Upper Class (20CU) in the sociological articulation constitutes a sub-profile since it is a profile of a sub-set of the articulation, namely, age, education, and social class. Other examples of sub-profiles are: Submissive-Low Participator (SL); 21 to 25-High School-Middle Class (25HM); Too Friendly (T).

In analogy with the *profile*, the *sub-profile* is the name of the descriptive category. *Cell* is the name given to the set of individuals characterized by a sub-profile. The cell has the same relation to the sub-profile as the unit cell has to the profile. In the above example all of the officer candidates who are classified 20CU would constitute the cell corresponding to that sub-profile. The cell would consist of all the unit cells corresponding to profiles that had the qualities 20, college, and upper class, regardless of their qualities in the other categories, hence the cell 20CU consists of the unit cells 20CUA and 20CUN.

Elements of the Calculus

A basic notion to the development of the calculus of empirical classes is that of the *indicator*. The *indicator* is defined as a function which coordinates to every profile some real number between 0 and 1, inclusive. Thus to every profile (or point in property-space) there corresponds only one real number in the unit interval (between 0 and 1, inclusive), but for a given number in this interval there may correspond many profiles. For example, the profile DFH may have assigned the number 0.61; DFL, 0.33; SFH, 0.61; STL, 0.94; etc., with each profile having exactly one number corresponding to it but a given number (e.g., 0.61) may have many profiles in correspondence.

The definition given for indicator illustrates the formal nature of this logical development. Indicator is defined strictly with reference to logical terms such as function, coordinates, number, with no commitment to empirical interpretation. As will be seen, an empirical meaning will later be suggested.

Every indicator determines a *class*, where class means the set of couples (each profile and the number assigned to it by the indicator) determined by the indicator. In the calculus of sets, a class may be defined as all of those individuals with a given property, say red. We may state this in CEC terms as the set of couples consisting of each red object and the number 1 assigned to it by an indicator function and each non-red object in the field and the corresponding number 0. The number 1 is taken to mean possession of the property red. The CEC concept of "class" would make this a special case of the more general case where each object has to a certain degree the property red, so, for example, a scarlet object would be assigned a higher number by the class indicator than an orange object, and much higher than a yellow object.

The number an indicator assigns a profile is called the *weight* from the profile to the class specified by the indicator. To illustrate, weights are assigned to all possible profiles for the psychological articulation of the field of inquiry (see Table 2).

There are 24 possible profiles. For an example of a class consider future army officers (Officer). Define Officer as "candidate who successfully completes officer training." Each cell is a profile, and each cell entry the weight of the profile to the class Officer. The indicator is the function which assigns a number to each profile, i.e., Table 2.

It is important to note the relation between a category, a quality, and a class. The differences seem to be explicable primarily in terms of whether or not the class is applied to an articulated or nonarticulated field. For example, suppose a field is articulated by the use of the one category Participation (qualities: High Participator, Low Participator) and then we apply the class Participation to the articulation. High Participator could

TABLE 2

Profiles and Hypothetical Weights to the Class "Officer" for the Psychological Articulation

	Dominance					
	Dominant		Independent		Submissive	
Friendliness	Participation		Participation		Participation	
	High	Low	High	Low	High	Low
Too Friendly	.29	.18	.43	.33	.20	.11
Friendly	.38	.31	.89	.14	.87	.14
Superficially Friendly	.47	.82	.29	.63	.44	.38
Not Friendly	.61	.19	.22	.06	.09	.52

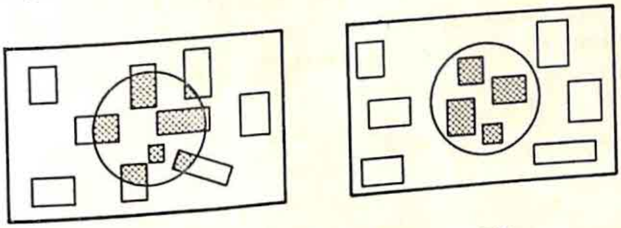
then be defined by giving a profile a weight of 0 if it were Low Participator and 1 if it were High Participator. However, if the class Popular is used, the weights from each profile (High Participator and Low Participator) would vary between 0 and 1, since participation and popularity are not perfectly related.

It is obvious that the situation could be reversed by articulating the field with the one *category* Popularity (qualities: Popular, Nonpopular), then apply Popular as a class to this new articulation and assign weights of 0 or 1 to every profile. Then the indicator specifying the class High Participator would assign weights from each profile *between* 0 and 1. Thus classes are very similar to qualities. The selection of a category and its qualities represents an act of choice in which all profiles by definition have one and only one quality from each category. Once the field has been articulated and a class introduced the weight from each profile is determined by its indicator function. Hence a quality could be considered a class to which each profile has the weight 0 or 1.

In social science there are similar terms. A category used to articulate the field is often called an independent variable and classes to which the individuals are referred are dependent variables. For example, independent variables may be religion of respondent which has the qualities, Protestant, Catholic, Jew, and income, with the qualities, High, Low. The dependent variable "votes Democratic" may then be introduced. A weight would be assigned from each profile to the dependent variable. The terms antecedent and consequent are sometimes used in the sense of category and class. Also observable and nonobservable variables, predictor and criterion variables, and manifest and latent variables are similar, respectively, to category and class.

Hence, in the example, Officer was a criterion or dependent variable, but an antecedent explanatory variable like father's ambition as the class could have been chosen equally well. Similarly, latent variables such as anxiety could be used. The two cardinal points are (*i*) there is no restriction on type of variable included under class or category, (*ii*) the relation between class and category (or quality) is very close, the difference resting primarily on the choice of whether the class is applied to an articulated or unarticulated field.

An individual x is an *instance to the degree* p of a class, if the indicator specifying the class assigns to the profile of x the number p. Thus the CEC treatment differs from the set-theoretical explication of classes primarily in treating every individual as an instance to some degree or other of the class rather than taking some individuals to be instances without qualification and others not at all. Classes differ from one another in the degree of membership in them assigned to the various individuals in the field of inquiry. This fundamental notion of a class, as opposed to a set, may be represented diagrammatically. In Figure 1, a circle represents a set or class and rectangles represent individuals. The shaded portion represents the degree of membership of an individual in a class; for sets this degree is either 0 or 1, for classes it varies from 0 to 1. Thus the set is a special case of a class.



EMPIRICAL CLASS SET

FIGURE 1

Diagrammatic Representation of Empirical Class and Traditional Set

Operations on Classes

Operations on classes are defined by rules for the combinations of weights. Analogous to this are the rules for combining classes in the set calculus, such as product, sum, etc. In the set calculus definitions utilize membership or nonmembership in a class or combination of classes. If by contrast individuals are regarded as having membership to some degree in every class there must be a reformulation of class operations in terms of combinations of weights.

The guiding principles in formulating the rules for the combinations of weights are: (*i*) classes for which the weights are all 0 or 1 are to have the properties of sets; (*ii*) the class connectives (sum, product, etc.) in these cases are to reduce to the familiar set connectives; (*iii*) the weights are to correspond as closely as possible to probabilities. These conditions lead to interesting field characteristics. Their relation to cumulative or Guttman scales is discussed below.

A note on terminology: K-S use ρ to designate articulations but here numbers will be used in the service of simplicity. λ is the symbol for profile, and small letters a, b, c designate classes. The notation

$$W_1(\lambda, a) = p \tag{1}$$

states that the weight in articulation 1 from profile λ to class a is p. The notation K-S use (and followed here) is that of Reichenbach [5] and is read: the weight from λ to a is p, or, the weight of a, given λ. An example: if λ is the profile 25HLN and a is the class Officer, and articulation 2 is the sociological articulation, then

$$W_2(\lambda, a) = p \tag{2}$$

may be read "the weight from the profile 25HLN to the class Officer is p" or, using the probability interpretation of weight, "the probability that an officer candidate of profile 25HLN will become an army officer is equal to p." Feller's [1] equivalent notation is

$$W_2[a|\lambda] = p. \tag{3}$$

Class Product

The definition of *class product* will be given first since it is presupposed in the definitions of most of the other operations. Here, as in the conventional calculus, the product of two classes refers to the joint or overlapping class of individuals who have membership in both of the original classes simultaneously, e.g., the product of the classes Officer and Popular refers to all those individuals who are popular officers.

In the new calculus of empirical classes *class product* is defined as follows

$$\{c = a \cap b\} =_{\text{Df}} \{W(\lambda, c) = \min\,[W(\lambda, a), W(\lambda, b)]\}, \quad \text{for all } \lambda. \tag{4}$$

Equation (4) may be read "the weight from λ to class c, where c is the joint class of a and b, is equal to the lesser of the two weights (those from λ to a and from λ to b)."

The product of two classes is defined as that class whose weight from each profile is the lesser of the weights to the original classes from that profile. Suppose the weights from every profile to two different classes are known, e.g., Officers and Popular, in other words one knows the weight (or probability) of every possible type of candidate (profile) becoming an army officer, and also of being popular among the other candidates. The question is, how can one compute his probability of being both an army officer and popular.

In the set calculus this probability would, in general, be equal to the proportion of candidates who were both officers and popular, in other words, the proportion of candidates who had weight equal 1 for both the classes Officers and Popular. An individual case is immediately determinable as either both weights 1 or not. In CEC the weight of a candidate being both an officer and popular is equal to the lesser weight of the individual classes. If his chances of being popular are only 0.2, while his chances (interpreted weight) of being an officer are 0.7, his likelihood of being both an officer and popular is only 0.2. Here it becomes clear that weight cannot in general be given the meaning of empirical probability for this definition is very different from the familiar rule for the probability of the occurrence of a joint class which is the product of the individual probabilities assuming the classes are independent.

The only condition under which the above definition of the class product as a minimum corresponds to the familiar rules for empirical probabilities of a joint class is when one class is included in another. Consider the interpretation of weight as an empirical probability, i.e., the limit of relative frequency. If the properties Officer and Popular are independent, the probability of an individual being both popular and an officer is 0.14 (0.7 × 0.2), considerably less than the 0.2 called for by the CEC definition of class product. The CEC definition, when related to the probability interpretation, posits a certain degree of dependence between classes, specifically, that one includes the other. In other words, the only conditions under which being a popular officer has probability 0.2 is when every popular candidate also becomes an officer, i.e., when the probability of a popular candidate being an officer is 1.0. If this definition were extended to more classes such that the probability of an individual being in all n classes exactly equalled its least probability for any class, the n classes form a cumulative or Guttman scale. Hence, the interpretation of weight as an empirical probability is subject to considerable restriction.

This definition of class product was chosen for specific reasons, but before these can be outlined it is necessary to introduce a second operation.

Class Inclusion

The basic operation determining how the other operations should be defined is class inclusion. This operation defines the conditions under which

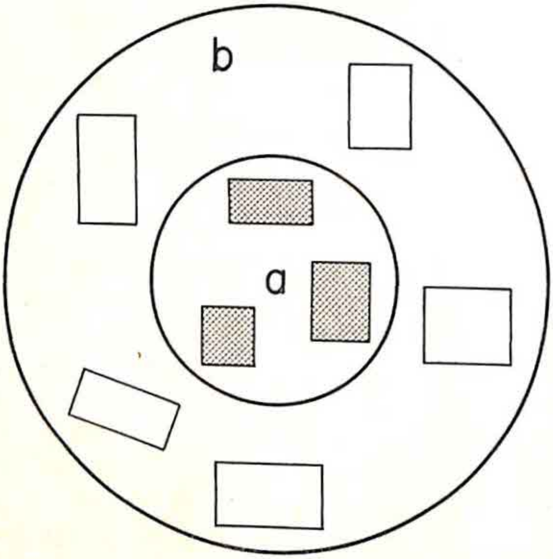


FIGURE 2
Class Inclusion in Set Calculus (a included in b)

one class is to be included within the other. In the regular set calculus inclusion is defined,

$$a \subset b =_{\mathrm{Df}} a = a \cap b, \tag{5}$$

which is read "the class a is included in the class b," equals by definition, "the class a is equal to the product (joint class) of a and b." Figure 2 represents this diagrammatically.

The problem then becomes one of defining inclusion in the calculus of empirical classes, analogous to the definition in the set calculus and such that it is equivalent to the set calculus in the extreme (i.e., when all weights equal zero or one). This was solved by making the conditions for class inclusion that, if a were included in b, the weight from every individual to a must be less than or equal to the weight from that individual to b. Figure 3 represents this diagrammatically. In this diagram the weight (proportion) of each individual must be less (or equal) to a than to b since all weights to a are also to b.

Hence the definition: Class Inclusion,

$$a \subset b =_{\mathrm{Df}} [W(\lambda, a) = W(\lambda, a \cap b)], \quad \text{for all} \quad \lambda. \tag{6}$$

Since class product is defined in terms of lesser weights from each profile, a is included in b if and only if the weight to a from each profile is less than or equal to that to b.

To follow our example, consider the classes Officer and Becomes Regular

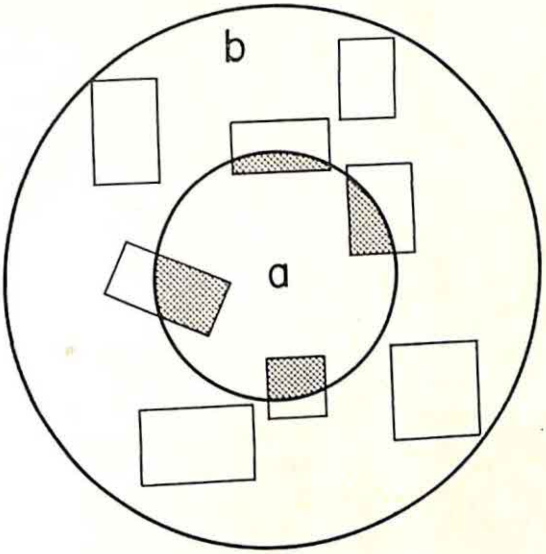


FIGURE 3
Class Inclusion in Empirical Class Calculus (a included in b)

Army Man (Regular). Officer includes Regular if for every profile it is more probable that a candidate with that profile becomes an officer than it is that he becomes a regular.

In the familiar set calculus, a is included in b means, in effect, "If an individual is a member of a he will surely be a member of b." In the new CEC, a is included in b is made to mean, in effect, "Insofar as an individual is a member of a he will at least as surely be a member of b." In more precise language, the combined result of the definition of class inclusion and class product is that if a is included in b, the weight from every individual to a must be less than or equal to the weight from that individual to b.

This, indeed, is the reason why the particular definition of class product was adopted in the CEC. For, as a result, the set calculus notion that "Regular is included in Officer if every man that is a Regular is also an Officer" is true in the ideal case of the empirical class calculus where all the weights from candidate to Officer are 1 and some from candidate to Regular are 1 and some are zero.

Class Sum

$$(7) \qquad [c = a \cup b] =_{\mathrm{Df}} [W(\lambda, c) = W(\lambda, a) + W(\lambda, b) - W(\lambda, a \cap b)]$$

for all λ.

The weight from each profile to the sum of two classes is defined (as in the probability calculus) as the sum of the weights diminished by the weight of the class product. By the definitions of class product, this is equivalent to taking the greater of the original weights

(8) $$[W(\lambda, a) \quad \text{or} \quad W(\lambda, b)].$$

In the set calculus the class sum is simply all the individuals who are members of (have weight 1 to) either class. This is easily computed by adding the individuals who belong to each class and subtracting those who are members of both classes (and therefore who have been counted twice). By analogy, the empirical class calculus would say that the weight of an individual being either an Officer or Popular is the sum of the weight for being an Officer and for being Popular minus the weight of being both. But since being both (the class product) has a weight equal to the lesser of the two original weights, the class sum becomes: greater weight plus lesser weight minus lesser weight. Thus the weight for being either an officer or popular is equal to whichever weight is higher. The class sum is then defined as that class whose weight from each profile is the greater of the weights to the original classes to that profile.

Class Contrary

(9) $$[b = -a] =_{\text{Df}} [W(\lambda, b) + W(\lambda, a) = 1], \quad \text{for all} \quad \lambda.$$

The contrary of a class is defined as the class to which the weight from each profile differs from unity by the weight to the original class. The interesting property of this definition is that it does not satisfy the law of the excluded middle (i.e., the law which states that an individual cannot belong to both a class and its contrary), since the weight from each profile to the class sum of a class and its contrary is only the larger of the two weights, not their sum (which is one). As the weights to a class approach the extreme values of 0 and 1, the weights to the sum of the class and its contrary approach 1, until in the limiting case the law of the excluded middle is satisfied.

Class Difference

In the set calculus the difference between two sets is definable in terms of the set complement, i.e.,

(10) $$(\alpha - \rho) =_{\text{Df}} (\alpha \cap - \rho)$$

(α minus ρ equals by definition α and not ρ). However, the corresponding formula in terms of the contrary of a class, $a \cap -b$, is not interpretable in general as a but not b since its weight from each profile is either that of a or of the contrary b, whichever is smaller. Hence *class difference* is defined as that class whose weight from any profile is equal to the difference in the weights from that profile to the original classes, respectively, except that it is 0 if that difference is not positive.

(11) $\{c = a - b\} =_{\mathrm{Df}} \{W(\lambda, c) = \max\,[W(\lambda, a) - W(\lambda, b), 0]\}$, for all λ.

In the limiting case where the class b has weights of only 0 or 1, the class contrary is directly analogous to the set complement, and, as in the set calculus, $a - b = a \cap -b$.

To illustrate the above operations for the psychological articulation, hypothetical values of the weights from each profile to the class sum, class product, class contrary, and class inclusion will be computed. Only five profiles will be used since they shall suffice for this example, (see Table 3).

TABLE 3

Hypothetical Values for Profile Weights to Various Class Operations

			Product	Sum	Contrary	Difference
Profile	$W(\lambda_1, a)$	$W(\lambda_1, b)$	$W(\lambda_1, a \cap b)$	$W(\lambda_1, a \cup b)$	$W(\lambda_1, -a)$	$W(\lambda, a-b)$
λ 1	0.89	0.81	0.81	0.89	0.11	0.08
λ 2	0.61	0.03	0.03	0.61	0.39	0.58
λ 3	0.53	0.72	0.53	0.72	0.47	0.00
λ 4	0.32	0.18	0.18	0.32	0.68	0.14
λ 5	0.12	0.12	0.12	0.12	0.88	0.00

An example of class inclusion is provided by the product and sum where the product is included in the sum, since for every profile the weight to the product is less than (or equal to) the weight to the sum.

Class product and class difference explicate the application of this calculus to combinations of responses on a questionnaire. For example, assume membership in an empirical class is interpreted as giving a positive response to an item, and weight from a profile to a class is interpreted as the nominal probability (to distinguish it from the relative frequency probability) of an individual or type (the profile) giving a positive response to a given item (the class). Then the nominal probability of a positive response on two items $(++)$ is the class product of the two weights, while the nominal probability of a positive response on the first item and a negative response on the second item $(+-)$ is equal to the class difference. Thus from the product and the difference the nominal probabilities of all possible response patterns could be computed. Again the isomorphism with the cumulative scale should be noted.

This completes the fundamental definitions which allow for correspondence (in the extreme) of this calculus for empirical classes to the usual set calculus. The purpose of the calculus so far is to transform the dichotomy of membership-nonmembership, basic to the set calculus, into a system with the fundamental properties and operations of the set calculus but incorporating the notion of weights or probabilities to the relation between an individual and a class. This makes the set calculus a special case of the class calculus where the probability of any individual belonging to any class is either 0 or 1.

Definitions of the four operations allow the development of a calculus of classes which corresponds in many respects to the usual calculus of sets. The following set calculus laws hold in this class calculus (with appropriate substitution of symbols): the laws of commutativity and associativity for sum and product, DeMorgan's theorems, the laws of tautology, the two distribution laws, the laws of double negation and of transposition and the transitivity of inclusion.

One more concept is needed to complete the analogue to the set calculus. Define *membership to the degree p*, to be symbolized by ϵ^p, as the relation of an individual to a class when the weight from the profile characterizing the individual to the class is p. Class analogues can then be given for many propositions, for example, the principle of extensionality becomes,

$$[a = b] = [X\epsilon^p a = X\epsilon^p b], \quad \text{for all} \quad X. \tag{12}$$

In other words, for two classes to be equivalent, all individuals must have membership to the same degree in each class. In our example Army Officers and Graduated Candidate School would in most cases be equivalent classes.

K-S then proceed to develop this calculus to cover the topics of nominal and empirical probabilities, the reference function (the reverse of an indicator, namely the nominal probability that if a certain class exists, a certain profile will also exist), classes and correlations, and constructs and intervening variables. Space does not permit a detailed exposition of these topics but the reader is referred to the original article for a very stimulating development.

An Application of the Calculus of Empirical Classes to Latent Structure Analysis

It appears that latent structure analysis (LSA) pursues the same objectives as CEC, although the problem is approached quite differently. A comparison of one of the results of these two approaches may prove profitable. Since both CEC and LSA are rather elaborate theoretical structures, the possible points of comparison are very numerous. One aspect will be selected for exploration—the implications of CEC for the assumption of local independence in LSA. This is selected because (*i*) the local independence assumption is the fundamental one in LSA, (*ii*) this point will illustrate the value of comparing the approaches, and (*iii*) the result of this particular analysis may be of substantive interest.

CEC and Local Independence

Using one interpretation of CEC, an alternative assumption to local independence is suggested. This assumption, called the assumption of a *local scale*, has the advantage that it allows the items and latent classes of LSA to be interpreted as classes and profiles, respectively, in CEC and therefore falls heir to all of the logical power of the set calculus. The implications of this alternative assumption will be explored.

A Latent Structure Interpretation of the Calculus of Empirical Classes

Of the possible interpretations that could be made of the CEC, profiles will be interpreted as latent classes, and classes will be interpreted as manifest items or observations. Of course, other interpretations of this strictly logical calculus are possible, e.g., a reversal of the above. This one, however, seems to have some interesting implications. The interpretation construes individuals as basically divisible into latent classes (profiles) which are usually not directly measured but only approximated by inference made from observable or manifest data (CEC classes).

An interesting consequence of this interpretation is that latent probabilities become isomorphic with weights (assigned by the indicator function). This is convenient because the latent probabilities give the probabilities of a member of a given latent class accepting a given indicator. This is the same meaning the weights have in the CEC if profiles represent latent classes and classes represent items. Similarly, the recruitment probabilities of LSA correspond to the reference probabilities of CEC—both being the probability of a certain latent class (or profile) given the item response (or class). An example should make this clear.

For this illustration we shall construct a very complex example. As indicated in Table 4, assume eight latent classes (profiles) taken from our psychological articulation, and three dichotomous items as shown in Table 4. The cell entries present the probability that an individual with the given profile answers items a, b, and c, plus.

This same data is represented in standard LSA form, in Table 5. The entries in CEC terms are $W\ (\lambda, a)$ where λ_i takes on values of all profiles, or latent classes. R is the reference function.

To compute the recruitment (LSA) or reference (CEC) probabilities divide the weight by the total probability a given item. These values are $R\ (a, \lambda_i)$ and are given in Table 5.

Class Product and Class Sum

Using this interpretation one divergence between LSA and CEC appears with regard to the assumption regarding combination of probabilities. In

TABLE 4

Hypothetical Probabilities (Weights) from Latent Classes (Profiles) to Items (Classes)

	Item a					Item b					Item c			
	O		C			O		C			O		C	
	H	L	H	L		H	L	H	L		H	L	H	L
F	.8	.1	.3	.9	F	.3	.6	.8	.2	F	.1	.3	.4	.8
N	.7	.2	.5	.5	N	.1	.5	.6	.9	N	.2	.3	.2	.7

TABLE 5

Latent Class Form of Table 4

Profiles	$W(\lambda_1,a)$	$W(\lambda_1,b)$	$W(\lambda_1,c)$	$R(a,\lambda)$	$R(b,\lambda)$	$R(c,\lambda)$
Latent Class	p^1_a	p^1_b	p^1_c	r^a_1	r^b_1	r^c_1
OHF	.8	.3	.1	8/40	3/40	1/30
OHN	.7	.1	.2	7/40	1/40	2/30
OLF	.1	.6	.3	1/40	6/40	3/30
OLN	.2	.5	.3	2/40	5/40	3/30
CHF	.3	.8	.4	3/40	8/40	4/30
CHN	.5	.6	.2	5/40	6/40	2/30
CLF	.9	.2	.8	9/40	2/40	8/30
CLN	.5	.9	.7	5/40	9/40	7/30
Sum	4.0	4.0	3.0	1.0	1.0	1.0

TABLE 6

Probabilities of Each Response Pattern Computed by LSA and CEC

Latent Classes	+ +		+ -		- +		- -	
	CEC	(LSA)	CEC	(LSA)	CEC	(LSA)	CEC	(LSA)
OHF	.30	.24	.50	.56	.00	.06	.20	.14
OHN	.10	.07	.60	.63	.00	.03	.30	.27
OLF	.10	.06	.00	.04	.50	.54	.40	.36
OLN	.20	.10	.00	.10	.30	.40	.50	.40
CHF	.30	.24	.00	.06	.50	.56	.20	.14
CHN	.50	.30	.00	.20	.10	.30	.40	.20
CLF	.20	.18	.70	.72	.00	.02	.10	.08
CLN	.50	.45	.00	.05	.40	.45	.10	.05

the class product, for example, which becomes isomorphic with LSA response patterns, consider the case of the response pattern ++ for the latent class OHF. LSA would say the probability equals the product of the individual probabilities or .8(.3) = .24, while CEC would say that probability equals the lesser of the two individual probabilities, or .3. The probability of response pattern +++ for CEC = .100, for LSA = .024. In virtually every case the predictions would differ. Similarly, for class sum where LSA would say it is .8 + .3 − .8(.3) = .86 (for ++), CEC equates the probability of having + on *either* item *a* or *b* to the larger, namely, .8. For +++, LSA gives .8 + .3 + .1 − .24 − .03 − .03 + 0.24 = .874, while CEC gives simply .8, the largest value.

For class difference, *a* and not *b* (response pattern +−) there is also a discrepancy. CEC says that the probability of answering + to *a*, and − to *b* is equal to either the difference between *W* (λ, *a*) and *W* (λ, *b*) or zero, whichever is larger, while LSA treats this like any other class product and would simply take the product of *W* (λ, *a*) and 1 − *W* (λ, *b*). Thus, for example, for latent class CLF, the probability of response pattern +− for CEC equals .9 − .2 = .7, while LSA puts this probability of +− at .9(.8) = .72. For CHN, +− would give by CEC .5 − .6 = 0. Therefore, the probability of +− is zero, while for LSA it equals .5(.4) = .20.

With the notions of class product and class difference it is possible to compute all values from the CEC scheme of the probabilities of each response pattern for each profile. In Table 6 we do this for the first two items and contrast these values with those obtained by LSA.

The Local Scale Assumption

From the above it is clear that the CEC derived model differs from the LSA model in that it does not assume local independence. It will be recalled that the assumption of local independence of LSA states that for every latent class X, and every pair of items i and j

$$P^x_{ijk\cdots} = P^x_i P^x_j P^x_k \cdots . \tag{13}$$

The assumption being made in the CEC derived model is that

$$P^x_{ij\cdots} = \min\,[P^x_i\,,\, P^x_j \ldots], \tag{14}$$

$$P^x_{i\bar{j}\cdots} = \max\,[P^x_i - P^x_j\,,\, 0]. \tag{15}$$

That is, for any two items i and j, for members of a given latent class, the probability of accepting both items (++) equals the probability of accepting the lesser (rather than the product of the probabilities as in LSA), and the probability of accepting item i, and not accepting item j (+−) equals the difference in the two probabilities, or zero, whichever is larger. These two equations characterize the *local scale assumption*.

The reason for this name stems from the fact that this assumption is true when the items form a cumulative scale for the members of a latent class. That is, the circumstance under which the local scale assumption holds is that the members of a latent class are all perfect scale types (within usual error) on the set of items from which the latent classes are derived. For example, consider a five-point cumulative scale and assume the proportion of respondents in each scale type (i.e., having the same response pattern) is equal.

In Table 7 are the weights for each scale type for each item in a perfect scale.

TABLE 7

Weights for Each Scale Type for Each Item in a Perfect Scale

	Items				
Scale Types	A	B	C	D	E
+ + + + +	1	1	1	1	1
+ + + + -	1	1	1	1	0
+ + + - -	1	1	1	0	0
+ + - - -	1	1	0	0	0
+ - - - -	1	0	0	0	0
- - - - -	0	0	0	0	0
Mean	.83	.67	.50	.33	.17

To test the first assumption of a local scale,

$$P^{x}_{ij} = \min\,[P_i\,, P_j], \tag{16}$$

take any two items, say B and D. The local scale assumption would say the probability of this class accepting both B and D equals min [.67, .33] which is .33. Computing this probability for the above latent class,

$$(1 + 1 + 0 + 0 + 0 + 0)/6 = 2/6 = .33. \tag{17}$$

This result will be the same for any two items. To test the second assumption,

$$P^{x}_{ij} = \max\,[P_i - P_j\,, 0], \tag{18}$$

let us take any two items, say A and E. The probability of accepting A and rejecting E, according to the local scale assumption, equals max [.83 − .17, 0], which is .66. The computation of this probability for the above latent class is

$$(0 + 1 + 1 + 1 + 1 + 0)/6 = 4/6 = .66. \tag{19}$$

The zero term in the equation covers the case of nonscale types, that is the probability of + on C and − on B, a −+ response, is negative, therefore zero. This is the formal statement that there are no nonscale types, or that the class includes only those individuals who are scale types. Hence the scale-type latent class fulfills the requirements of the local scale assumption. It is likely that this is the only type of model that does, though that proof has not yet been worked out.

This assumption, which leads to a set of accounting equations, will not be pursued. The suggestion has been made by Lazarsfeld that this development could be pursued by studying the distribution of error. All that can be said at this time is that this appears to be a promising lead. The main point to be stressed here however is that LSA is based on the local independence assumption, but the local scale assumption and in general the consideration of other assumptions might render LSA more flexible and lead to different latent classes that may be more appropriate in some investigations.

REFERENCES

[1] Feller, W. *An introduction to probability and its applications.* New York: Wiley, 1950.
[2] Kaplan, A. Definition and specification of meaning. *J. Phil.*, 1946, **43**, 281-288. [Also reprinted in P. Lazarsfeld and M. Rosenberg (Eds.), *The Language of Social Research.* Glencoe, Ill.: Free Press, 1955. Pp. 527-532.]
[3] Kaplan, A. and Schott, H. F. A calculus for empirical classes. *Methodos*, 1951, **3**, 165-190.
[4] Lazarsfeld, P. and Barton, A. H. Qualitative measurement in the social sciences: classification, typologies, and indices. In D. Lerner and H. D. Lasswell (Eds.), *The policy sciences.* Stanford: Stanford Univ. Press, 1951. Pp. 155-192.
[5] Reichenbach, H. *The theory of probability.* Berkeley: Univ. California Press, 1949.

Manuscript received 1/10/58

Revised manuscript received 5/19/58

PSYCHOMETRIKA—VOL. 24, NO. 1
MARCH, 1959

NOTE ON $\phi/\phi_{\max}$

EDWARD E. CURETON

UNIVERSITY OF TENNESSEE

Formulas are given for a descriptive statistic related to the fourfold-point correlation but having always-attainable limits ± 1.

The fourfold-point correlation coefficient, ϕ, is derived from a table such as

	$-$	X	$+$	
$+$	$p_1 - p_{12}$		p_{12}	p_1
Y				
$-$	$q_2 - p_1 + p_{12}$		$p_2 - p_{12}$	q_1
	q_2		p_2	1

where p_1, q_1, p_2, and q_2 are the marginal proportions, and p_{12} is the joint positive proportion. Then

$$\phi = \frac{p_{12} - p_1 p_2}{\sqrt{p_1 q_1 p_2 q_2}} \tag{1}$$

is the product-moment correlation between the dichotomous varieties X and Y.

It is well known that ϕ can equal $+1$ only if $p_1 = p_2 = .5$, and that it can equal -1 only if $p_1 = q_2 = .5$ ([1], p. 324: [2], p. 342). Hence ϕ has limits ± 1 only if all marginal proportions are .5.

In order to obtain a descriptive statistic having always-attainable limits ± 1, it has been proposed that one use $\phi/\phi_{\max}$, $\phi_{\max}$ being the maximum value of ϕ consistent with the given marginal proportions. It is apparent from the table that the maximum value of p_{12} is p_1 or p_2, whichever is smaller (call it p'), and since the denominators of ϕ and $\phi_{\max}$ contain only the marginal proportions and must therefore be identical,

$$\frac{\phi}{\phi_{\max}} = \frac{p_{12} - p_1 p_2}{p' - p_1 p_2}. \tag{2}$$

This formula applies if ϕ is positive. But for ϕ negative, $\phi_{\max}$ should be the absolute value of the largest negative correlation consistent with the given marginals; this will occur when either p_{12} or $q_2 - p_1 + p_{12}$ is 0, and the denominator of $\phi/\phi_{\max}$ is then $p_1\, p_2$ minus the smallest possible value of p_{12}.

If p_{12} can take the value 0, the denominator is simply p_1p_2 . If p_{12} cannot take the value 0, $q_2 - p_1 + p_{12}$ can, in which case the smallest possible value of p_{12} is $p_1 - q_2$. For ϕ negative, therefore,

$$\frac{\phi}{\phi_{\max}} = \frac{p_{12} - p_1p_2}{p_1p_2 - (p_1 - q_2)}, \tag{3}$$

and the second term in the denominator is subtracted from p_1p_2 only if $p_1 > q_2$.

More convenient computing formulas are obtained if one multiplies numerator and denominator of (2) and (3) by N^2:

$$\frac{\phi}{\phi_{\max}} = \frac{NS_{12} - S_1S_2}{NS' - S_1S_2}, \phi \text{ positive;} \tag{4}$$

$$\frac{\phi}{\phi_{\max}} = \frac{NS_{12} - S_1S_2}{S_1S_2 - N(S_1 + S_2 - N)}, \phi \text{ negative.} \tag{5}$$

Here S_1 and S_2 are the positive marginal frequencies, S_{12} the joint positive frequency, $S' = S_1$ or S_2 , whichever is smaller, and the second term of the denominator of (5) is subtracted from S_1S_2 only if $S_1 + S_2 > N$.

For the alternative table,*

		X	
	$-$	$+$	
$+$	a	b	S_1
Y			
$-$	c	d	$N - S_1$
	$N - S_2$	S_2	N

the numerators of (4) and (5) are equal to $bc - ad$. For ϕ positive, either a or d must be 0 when $\phi = \phi_{\max}$, so

$$\frac{\phi}{\phi_{\max}} = \frac{bc - ad}{b'c'}, \phi \text{ positive,} \tag{6}$$

and b' and c' are obtained from a fourfold having the same marginals but with a or d (whichever is smaller) replaced by 0. For ϕ negative, either b or c must be 0 when ϕ takes its maximum negative value, so

$$\frac{\phi}{\phi_{\max}} = \frac{bc - ad}{a'\,d'}, \phi \text{ negative,} \tag{7}$$

and a' and d' are obtained from a fourfold having the same marginals but with b or c (whichever is smaller) replaced by 0.

The limitation of range of the ϕ coefficient is a special case. The product-moment correlation always has attainable limits ± 1 if both variates are continuous, or more generally if no two individuals have identical scores on either variate. With grouped or discrete distributions, its upper limit is + 1 only if the two marginal distributions are identical, its lower limit is − 1 only if one marginal distribution is identical with the other reversed, and its two limits are ± 1 only if both marginal distributions are symmetrical and identical.

---

*This alternative development is based on an editorial suggestion.

REFERENCES

[1] Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, **6**, 323-329.

[2] Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.

Manuscript received 1/27/58

Revised manuscript received 6/3/58

BOOK REVIEW

Clyde H. Coombs and Richard C. Kao. *Nonmetric Factor Analysis.* Ann Arbor: University of Michigan Engineering Research Institute Bulletin No. 38, 1955. Pp. vii + 63.

Following a brief description of Coombs' theory of data, this slender technical monograph treats two classes of models for the resolution of behavior into components. These models, conjunctive-disjunctive and compensatory, are intended for scaling dichotomous monotone items collected by the method of single stimuli. The models are not competitive with Guttman scalogram analysis but are offered as multidimensional formulations which might be appropriate when unidimensional solutions are not obtained. A general axiomatic basis for the approach is presented, and theorems specific to each model are derived separately. Some very helpful interpretive material is interspersed throughout the mathematical development.

In the conjunctive model success on a task requires a certain minimum on *each* of the relevant dimensions; in the disjunctive model success requires a minimum on any *one* of the dimensions involved. The conjunctive and disjunctive models, although psychologically distinct, are shown to be formally equivalent and are treated simultaneously throughout the development. In the compensatory model, which underlies multiple-regression theory and multiple-factor analysis, a given performance may be achieved by many different combinations of components: an excess in one attribute may compensate for a shortage in another. The logic of the models, then, would call for the use of multiple cutting scores for a conjunctive domain and multiple regression for a compensatory domain.

The general model is constructed by representing both individuals and stimuli as vectors in a multidimensional space, called the genotypic space. An individual's response to a stimulus is treated formally as an order relation on the corresponding individual vector and stimulus vector. Theorems are derived to generate a calculus for recovering the underlying genotypic structure from the data. In the conjunctive model the measures in the genotypic space are assumed to be elements of only an ordinal scale, and the analysis recovers the several components at the level of partial orders or simple orders. A method is also described for estimating a lower bound of the dimensionality.

In the compensatory model the genotypic structure is strengthened to the level of real numbers, but the additional assumptions required to recover the elements at that level are not made, so that factor loadings are obtained on an ordinal scale. Both an individual compensatory and a stimulus compensatory model are constructed, depending on whether the linear function which weights the several components is characteristic of an individual and independent of the stimuli, or vice versa. Only the rudiments of the compensatory model have been formulated. Procedures have been developed only for the special case of two dimensions, and the information recovered in the analysis is the order of the factor loadings (cosine of the angle between vector and coordinate axis), which does not correspond to the order of the projections.

These scaling models face several problems which are anticipated by Coombs and Kao in their discussion section. The problem of error is not considered in these completely deterministic models, and no procedures are offered for its treatment. The authors are aware, however, that ". . . the data will contain error, and some sort of stochastic generalizations will have to be constructed to make these models useful. . . ." Another problem arises with respect to uniqueness in handling incomplete data; Milholland has found that alternative solutions can be obtained in the conjunctive-disjunctive model at least. Another important concern involves criteria for choosing between the models for application to data. The authors suggest that an appropriate model would require no more dimensions than an inappropriate one and would generate less nonexistent data.

Several advantages may accrue from a mathematically precise formulation of a measurement model, not the least of which is the clarity gained from an explicit statement of assumptions and inferences. For example, the precise formalization of a model might stimulate theory construction in measurement by providing the foundation for further generalizations and by making it possible to compare the exact properties of different methods. A more important contribution is sometimes made, however, in that the models underlying various measurement methods frequently have implications for theorizing in psychology. Some such implications, a few of which are briefly illustrated in the Introduction, are attendant upon the distinctions made by Coombs and Kao in formulating the models of *Nonmetric Factor Analysis*. These implications may be derived in spite of the fact that statistical machinery has not yet been developed for practical applications of the models.

The elegance and precision gained through the axiomatic presentation has an accompanying difficulty of reading, so that the audience for this technical monograph seems limited to measurement specialists interested in the theory of models.

Educational Testing Service SAMUEL MESSICK

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

ON METHODS IN THE ANALYSIS OF PROFILE DATA

SAMUEL W. GREENHOUSE AND SEYMOUR GEISSER*

NATIONAL INSTITUTE OF MENTAL HEALTH

This paper is concerned with methods for analyzing quantitative, noncategorical profile data, e.g., a battery of tests given to individuals in one or more groups. It is assumed that the variables have a multinormal distribution with an arbitrary variance-covariance matrix. Approximate procedures based on classical analysis of variance are presented, including an adjustment to the degrees of freedom resulting in conservative F tests. These can be applied to the case where the variance-covariance matrices differ from group to group. In addition, exact generalized multivariate analysis methods are discussed. Examples are given illustrating both techniques.

Much research in the social sciences is of the multivariate type; multiple observations are made on individuals who have been sampled from one or more populations. In particular, when the observations are in the form of a battery of tests or a set of items, there is the problem of profile analysis, wherein it is customary to test for differences in the levels and in the shapes of the group profiles. If the variables being observed are assigned to columns and the individuals to rows, the resulting matrix of observations is very suggestive of the data usually analyzed by analysis of variance. Furthermore, since the rows are random and the columns can be considered in almost all instances as fixed, the appropriate model is the mixed model.

As is well known, in order that the usually computed ratios of mean squares in this model [7, 14, 16] be exactly distributed as the F distribution, it is necessary that columns (variables), in addition to being normally distributed, have equal variances and be mutually independent or, at most, have equal correlations. But these assumptions seem much too restrictive. In most investigations, it is unrealistic to assume that three or more tests, items, or treatment schedules have the same pairwise correlations or that they have the same variances. It seemed obvious, therefore, that this problem of multiple observations should be considered in its greatest generality, namely, that an individual vector $x_1, x_2, \cdots, x_p$ is sampled from a p-variate normal distribution with an arbitrary variance-covariance matrix.

Exact procedures for analyzing data of this type have been known for some time and are usually referred to as the generalized multivariate analysis of variance [1, 10, 12, 13, 17]. These, however, require considerably more computations than that demanded by the arithmetic of the analysis of variance. Furthermore, an analysis of variance approach permits the analysis of a set of data which cannot be handled by multivariate procedures, namely, the case where n, the number of random vectors, is less than p, the number of variables. Although these multivariate methods are discussed subsequently and an example is given for the case of two groups, our main purpose is to utilize the simpler, and more familiar, conventional univariate analysis of variance techniques under the more general assumptions. Our results concerning the approximate distributions of the F statistics are based upon the work of Box [5, 6] with regard to one group and its extension, by Geisser and Greenhouse [8], to several groups. In addition, the latter have found certain adjustments to the approximate tests leading to conservative tests which can be used, when the group sample sizes are the same, in the case of unequal variance-covariance matrices among the groups.

---

*We are indebted to Mrs. Norma French for performing all the calculations appearing in this paper.

It is of interest that Block, Levine, and McNemar [2] were also primarily concerned with the application of the analysis of variance to the profile problem. They presented F tests for testing the homogeneity of variable (columns) means, the homogeneity of over-all group means (profile levels) and the equality of profile shapes. However, they assumed equal variance, among the variables and, since they imply that the F tests are exact, it can only be inferred that they also assumed the variables to be independent or equally correlated.

The Problem

Our notation is almost identical to that used by Block, Levine, and McNemar. Let p tests, $x_1, x_2, \cdots, x_p$, be given to each of n_k individuals ($k = 1, 2, \cdots, g$) in each of g established groups. Assume both the p tests and the g groups to be fixed, i.e., they are not random elements sampled from larger populations. This model, which fixes interest in the tests and groups under study, conforms to many experimental situations met with in practice. The totality of Np observed scores ($N = \sum_{k=1}^{g} n_k$) can be classified according to the scheme at the top of the next page.

An individual i in group k has the profile

$$(x_{i1k}, \cdots, x_{ijk}, \cdots, x_{ipk}) \quad \begin{cases} i = 1, \cdots, n_k & \text{individuals in group } k \\ j = 1, \cdots, p & \text{variables} \\ k = 1, \cdots, g & \text{groups.} \end{cases}$$

And the group profile for group k, say, is represented by

$$(\bar{x}_{.1k}, \bar{x}_{.2k}, \cdots, \bar{x}_{.pk}).$$

		Tests					
Group	Ind.	x_1	$\dots$	x_j	$\dots$	x_p	
1	1	x_{111}		x_{1j1}		x_{1p1}	
	$\vdots$						
	n_1	x_{n_111}		x_{n_1j1}		x_{n_1p1}	
Means: Gr. 1		$\bar{x}_{\cdot 11}$		$\bar{x}_{\cdot j1}$		$\bar{x}_{\cdot p1}$	$\bar{x}_{\cdot\cdot 1}$
	$\vdots$						
k	1	x_{11k}		x_{1jk}		x_{1pk}	
	$\vdots$						
	i	x_{i1k}		x_{ijk}		x_{ipk}	
	$\vdots$						
	n_k	x_{n_k1k}		x_{n_kjk}		x_{n_kpk}	
Means: Gr. k		$\bar{x}_{\cdot 1k}$		$\bar{x}_{\cdot jk}$		$\bar{x}_{\cdot pk}$	$\bar{x}_{\cdot\cdot k}$
	$\vdots$						
g	1	x_{11g}		x_{1jg}		x_{1pg}	
	$\vdots$						
	n_g	x_{n_g1g}		x_{n_gjg}		x_{n_gpg}	
Means: Gr. g		$\bar{x}_{\cdot 1g}$		$\bar{x}_{\cdot jg}$		$\bar{x}_{\cdot pg}$	$\bar{x}_{\cdot\cdot g}$
Means: All Groups		$\bar{x}_{\cdot 1\cdot}$		$\bar{x}_{\cdot j\cdot}$		$\bar{x}_{\cdot p\cdot}$	$\bar{x}_{\cdot\cdot\cdot}$

Assume that each individual profile is a random vector sampled from a p-variate normal distribution with an arbitrary variance-covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1p}\sigma_1\sigma_p \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2p}\sigma_2\sigma_p \\ \vdots & \vdots & & \vdots \\ \rho_{1p}\sigma_1\sigma_p & \rho_{2p}\sigma_2\sigma_p & \cdots & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Also assume that the p variables have the same metric. This is necessary to give meaning to the question of whether the group profiles have the same shape, and not because of any statistical considerations. This restriction results in no loss of generality if there already exists a large body of data on these p tests so that the standard deviations can be assumed known. For in this instance, equal metrics can be obtained by standardization of the p test scores.

The questions that are most often asked in profile analysis are:

(i) Are the groups on the same level, i.e., do the groups arise from populations having the same group means, namely, $E(\bar{x}_{..1}) = E(\bar{x}_{..2}) = \cdots = E(\bar{x}_{..g})$, where E denotes the expectation?

(ii) Do the groups have the same shape, i.e., do the groups arise from populations having parallel group profiles?

Another question that may be asked of these data, although not too frequently in profile analysis, is whether the p tests have the same means.

With regard to the question on shape, it becomes necessary to define a statistic which reflects the concept of equally shaped group profiles. In a larger sense, profiles having the same shape can be considered to be parallel curves. As Box [4] and Block, Levine, and McNemar [2] point out, parallelism can be measured by the group-test interaction mean square. That is, if the curves are parallel, the group-test interaction should be zero and the mean square should not differ significantly from an appropriate error mean square. If, on the other hand, the curves have different shapes, the interaction mean square should be significantly greater than the error mean square.

This is made clear by reference to two group profiles:

$$\bar{x}_{.11}, \bar{x}_{.21}, \cdots, \bar{x}_{.p1}$$

and

$$\bar{x}_{.12}, \bar{x}_{.22}, \cdots, \bar{x}_{.p2}.$$

Denote the corresponding differences between group means for each test by

$$d_1, d_2, \cdots, d_p.$$

If the two profiles are parallel it is clear that $d_1 = d_2 = d_3 = \cdots = d_p$. On the other hand, if $d_1 = d_2 = d_3 = \cdots = d_p$, then the two profiles must be parallel. Hence, a necessary and sufficient condition that the two group profiles possess the same shape is that

$$d_1 = d_2 = d_3 = \cdots = d_p.$$

But the equality of these differences is exactly what is meant by no interaction between groups and tests, and the extent to which these differences are unequal corresponds to the existence of the group-test interaction. Therefore a test of the group-test interaction is also a test of whether group profiles have the same shape.

Tests of Significance In the Mixed Model

If the p test scores have equal variances and are independent (or, at most, are equally correlated in pairs), so that

$$\Sigma_0 = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

or

$$\Sigma_1 = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

then the given scheme constitutes the classical mixed model for g samples, with proportionate numbers of observations among the samples. The appropriate analysis of variance breakdown is shown in Table 1. The analysis under either of the above assumptions on the covariance matrix follows along classical lines. The F_1, F_2, and F_3 statistics used to test hypotheses of homogeneity of test (variable) means, of group means (level) and the nonexistence of a group-test interaction (equal shapes of group profiles), respectively, are exact.

If, on the other hand, the validity of these two models is suspect, on the basis either of prior evidence or of a statistical test, the given F ratios are not distributed like the tabulated F distribution. In this situation where the covariance matrix is assumed to be arbitrary and given by Σ, Roy [13], Rao [12], and others have approached the problem through the multivariate analysis of variance. However, it is of interest, and possibly of considerable practical importance, to investigate the distribution of the computed F statistics.

Tests of Significance for Arbitrary Covariance Matrix

Geisser and Greenhouse [8], in extending to several groups Box's work [5, 6] relative to one group, have shown that Q_1 and Q_4 are each independent of Q_5, and Q_2 is independent of Q_3. They have also shown that, under the null hypothesis,

$$E(Q_1) = A, \quad \text{say}, \quad E(Q_4) = (g-1)A, \qquad E(Q_5) = (N-g)A,$$

and

$$E(Q_2) = (g-1)B, \quad \text{say}, \quad E(Q_3) = (N-g)B.$$

Table 2 gives the mean square (M.S.) and the expectations of the mean

TABLE 1

Analysis of Variance

Source	d.f.	Sum of Squares	F
Tests	p-1	$Q_1 = N\sum_{j=1}^{p}(\bar{x}_{.j.} - \bar{x}_{...})^2$	$F_1 = (N-g)\,\frac{Q_1}{Q_5}$
Groups	g-1	$Q_2 = p\sum_{k=1}^{g} n_k(\bar{x}_{..k} - \bar{x}_{...})^2$	$F_2 = \frac{(N-g)\,Q_2}{(g-1)\,Q_3}$
Individuals (within Groups)	N-g	$Q_3 = p\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{x}_{i.k} - \bar{x}_{..k})^2$	
Group x Tests	(p-1)(g-1)	$Q_4 = \sum_{k=1}^{g}\sum_{j=1}^{p} n_k(\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...})^2$	$F_3 = \frac{(N-g)\,Q_4}{(g-1)\,Q_5}$
Indiv. x Tests (within Groups)	(p-1)(N-g)	$Q_5 = \sum_{k=1}^{g}\sum_{i=1}^{n_k}\sum_{j=1}^{p}(x_{ijk} - \bar{x}_{.jk} - \bar{x}_{i.k} - \bar{x}_{..k})^2$	
Total	Np-1	$Q_6 = \sum_{k}\sum_{i}\sum_{j}(x_{ijk} - \bar{x}_{...})^2$	

square (E.M.S.) for each of the five sources of variation in the analysis of variance.

From the results presented in Tables 1 and 2, it follows that each of the three F ratios, F_1, F_2, and F_3, is a ratio of two independent mean squares with the same expectations under the null hypothesis. Making use of the fact that each of the quadratic forms involved in the three F statistics is exactly distributed like a linear sum of independent χ^2 variables with the same degrees of freedom (theorem 6.1, Box [5, 6]), F_1 is approximately distributed like $F[(p-1)\epsilon, (p-1)(N-g)\epsilon]$, F_3 is approximately distributed like $F[(p-1)(g-1)\epsilon, (p-1)(N-g)\epsilon]$, and F_2 is exactly distributed like $F(g-1, N-g)$, where

TABLE 2

Analysis of Variance

Source	M.S.	E.M.S.
Tests	$(p-1)^{-1}Q_1$	$(p-1)^{-1}A$
Groups	$(g-1)^{-1}Q_2$	B
Individuals within Groups	$(N-g)^{-1}Q_3$	B
Groups x Tests	$(p-1)^{-1}(g-1)^{-1}Q_4$	$(p-1)^{-1}A$
Individuals x Tests within Groups	$(p-1)^{-1}(N-g)^{-1}Q_5$	$(p-1)^{-1}A$

$$\epsilon = p^2(\bar{\sigma}_{tt} - \bar{\sigma}_{..})^2/(p-1)(\Sigma\Sigma\sigma_{ts}^2 - 2p\Sigma\bar{\sigma}_{t.}^2 + p^2\bar{\sigma}_{..}^2);$$

σ_{ts} are the elements of the matrix Σ, $\bar{\sigma}_{tt}$ is the mean of the diagonal terms, $\bar{\sigma}_{t.}$ is the mean of the tth row (or tth column), and $\bar{\sigma}_{..}$ is the grand mean. Thus, the effect of the arbitrary variance-covariance matrix, which must be the same from group to group, is to assess the significance of the F_1 and F_3 statistics in the ordinary tabulated F distribution but with reduced degrees of freedom. The F_2 test on group means, it will be noted, remains unchanged from the standard F test since it results from a one-way analysis of variance with all observations having the same variance.

The reduction in the degrees of freedom for this approximate test is a function of the elements of the population variance-covariance matrix. This is almost never known, and therefore ϵ will have to be estimated from the sample variances and covariances. However, the effect of using an estimated ϵ on the approximate F distributions involved is unknown. Hence, unless the variance-covariance matrix is estimated with a large number of degrees of freedom, use of the conservative test given below is suggested.

A Conservative Test

The preceding approximate procedure requires some computations on the elements of a known variance-covariance matrix. In many profile problems, the number of tests may be as high as 50 if not more. This results in a 50 × 50 matrix, necessitating some laborious arithmetic. Furthermore, in almost all problems variances and covariances are unknown and the extent to which ϵ is changed by using sample estimates has not been investigated. As a result it is useful to obtain a lower bound on ϵ; it can be shown that

$$\epsilon > \frac{1}{p-1}.$$

This minimum value of ϵ is independent of the elements of the variance-covariance matrix.

With this new correction to the degrees of freedom, the F_1 and F_3 statistics are now judged for significance by entering the tabulated F distribution with 1 and $N - g$ degrees of freedom and with $g - 1$ and $N - g$ degrees of freedom respectively. These tests are called conservative since the minimum value of ϵ gives the maximum reduction in degress of freedom.

An Example

Five groups of mothers, classified into their groups according to some external criteria, were given a maternal attitude questionnaire containing 23 scales. For purposes of this illustration, six of these scales have been selected. Thus $p = 6$, $g = 5$, and $N = 128$. The group profiles and group means are given in Table 3.

The five variance-covariance matrices were first tested for homogeneity. The likelihood ratio test, the multivariate analogue of Bartlett's test for

TABLE 3

Mean Profiles for Five Groups of Mothers on Selected Scales of a Maternal Attitude Questionnaire*

Groups	No. of Mothers	Scale 1	Scale 3	Scale 6	Scale 9	Scale 13	Scale 14	Group Mean
A	59	17.02	10.97	13.24	11.47	9.80	15.44	12.99
B	13	17.92	13.85	17.23	14.00	12.23	17.38	15.44
C	15	18.87	11.60	14.13	8.93	8.27	17.73	13.26
D	32	16.75	14.47	15.41	11.78	9.91	15.94	14.04
E	9	18.33	10.78	13.89	14.44	12.11	18.78	14.72
All Groups	128	17.35	12.20	14.34	11.72	10.05	16.27	13.65

*We are indebted to Dr. Richard Q. Bell of the Laboratory of Psychology, National Institute of Mental Health, for permitting us to use part of his data for this example.

homogeneity of variances, can be found in Box [3, 4]. (Kullback [11] derives an equivalent test through information theory.) The test statistic is

$$M = N \log_e | S | - \sum_{i=1}^{5} n_i \log_e | S_i | = 112.6565.$$

In the above $| S |$ is the determinant of the pooled variance-covariance matrix, and $| S_i |$ is the determinant of the sample variance-covariance matrix in the ith group. Now compute

$$A_1 = \frac{2p^2 + 3p - 1}{6(p + 1)(g - 1)} \left(\sum_{i=1}^{g} \frac{1}{n_i} - \frac{1}{N} \right) = .17012,$$

and

$$f_1 = \tfrac{1}{2}p(p + 1)(g - 1) = 84,$$

and enter $(1 - A_1)M = 93.4$ in the χ^2 distribution with 84 degrees of freedom. Since the probability of getting this value of χ^2 or larger is fairly high, the null hypothesis of equal variance-covariance matrices is not rejected.

An estimate of the matrix Σ is given by the pooled variance-covariance matrix

$$S = \begin{bmatrix} 3.100 & .101 & -.279 & -.083 & -.009 & 1.557 \\ .101 & 5.780 & 1.013 & -.114 & -1.014 & .039 \\ -.279 & 1.013 & 5.560 & 1.039 & 1.366 & -.169 \\ -.083 & -.114 & 1.039 & 5.600 & 3.080 & .258 \\ -.009 & -1.014 & 1.366 & 3.080 & 6.820 & .222 \\ 1.557 & .039 & -.169 & .258 & .222 & 5.170 \end{bmatrix}.$$

Consider now whether the hypothesis of equal variances and equal covariances is consistent with S. The best estimate of the uniform variance-covariance matrix under this hypothesis is given by

$$S_1 = \begin{bmatrix} 5.3888 & .467 & \cdots & .467 \\ .467 & 5.338 & \cdots & .467 \\ \vdots & \vdots & & \vdots \\ .467 & .467 & \cdots & 5.338 \end{bmatrix},$$

where the diagonal element is an average of the 6 variances in S and the covariance is an average of the $15[\frac{1}{2}p(p - 1)]$ covariances in S. The reason for testing this hypothesis is that if S_1 is consistent with the data then classical analysis of variance procedures are applicable. The test used is again a likelihood ratio test, also given by Box [3, 4]. The test statistic is

$$M = -(N - g) \log_e \frac{|S|}{|S_1|} = -123 \log_e \frac{10{,}806.42}{21{,}040.49} = 81.956,$$

where $(N - g) = 123$ is the degrees of freedom entering into the computation of any element in S or S_1. Now compute

$$A_1 = \frac{p(p+1)^2(2p-3)}{6(N-g)(p-1)(p^2+p-4)} = .01887,$$

and

$$f_1 = (p^2 + p - 4)/2 = 19,$$

and enter $(1 - A_1)M = 80.4$ in the χ^2 tables with $f_1 = 19$ degrees of freedom. The probability of this result is well below .001; the hypothesis of equal variances and equal covariances must be rejected.

The analysis of variance yields the numerical results of Table 4.

TABLE 4

Analysis of Variance

Source	d.f.	SS	M.S.	F
Tests	5	5092.56		
Groups	4	509.12	127.28	$F_2 = 16.51$
Individuals within Groups	123	948.41	7.71	
Groups x Tests	20	644.74	32.24	$F_3 = 6.63$
Individuals x Tests within Groups	615	2991.04	4.86	

Of primary interest is the test of the homogeneity of group profiles, which is a test for the existence of the group-test interaction. For this purpose enter the F_3 value in the F table with $(g - 1)(p - 1)\epsilon$ and $(N - g)(p - 1)\epsilon$, or with 20ϵ and 615ϵ, degrees of freedom. From the previous formula, and the elements in the S matrix, ϵ is estimated to be .8194. Therefore the effective degrees of freedom are 16 and 503. The observed $F_3 = 6.63$ is greater than the .001 point for F with 15 and 120 degrees of freedom. One therefore rejects the hypothesis of no interaction and concludes that the mean profiles differ in shape from group to group.

The conservative test, which of course does not require the computation of ϵ, would enter $F_3 = 6.63$ in the F tables with $g - 1 = 4$ and $N - g = 123$ degrees of freedom. The .001 point for F with these degrees of freedom is 4.95. In this case, therefore, the conservative test yields the same conclusion as the approximate test, namely, the probability that the group profiles differ in shape due to chance is less than .001.

The groups clearly differ with regard to levels as can be seen from the very large F_2 value.

Other Procedures

The foregoing procedures present approximate and conservative tests of significance resulting from the analysis of variance utilizing readily available tables of the F distribution. As mentioned earlier there are available exact procedures in the multivariate analysis of variance. These procedures lead to exact tests of the general hypothesis in multivariate analysis of the equality of vector means among g populations and of the existence of the group-item or group-test interaction of interest in profile analysis. However, all of these procedures require laborious computations involving the inversion of $(p \times p)$ matrices (p equal to the number of tests or items) and the computation of latent roots or the evaluation of determinants. A further complication is the lack of tabled probability values for the appropriate test statistics. Recently, however, distribution tables have appeared relating to the approach of multivariate analysis initially taken by Roy [13]. Under this view, the distribution of the test statistic is dependent upon the distribution of the maximum characteristic root of certain matrices. The most comprehensive tables or charts thus far available are those given by Heck [9]. Heck, incidentally, specifically considers the problem of profile analysis.

The case for two groups will be developed in some detail to illustrate the principles involved and then the extension to g groups as given by Heck will be summarized briefly. The former situation leads to Hotelling's generalized T^2 statistic and is implied in the literature on multivariate analysis.

In the previous notation, x_{ijk} is an observation on item j for individual i in group k, and $\bar{x}_{.jk}$ is the mean of character j in group k. The range of subscripts here is $k = 1, 2$; $j = 1, 2, \cdots, p$; and $i = 1, 2, \cdots, n_k$. As before, assume that the random vector $x'_{i(k)} = (x_{i1k}, \cdots, x_{ipk})$ is $N(\mu_{(k)}, \Sigma)$, that is, the p variables have a multivariate normal distribution in population k with mean vector $\mu'_{(k)} = (\mu_{1k}, \cdots, \mu_{pk})$ and variance-covariance matrix Σ which is common to the g populations. The hypothesis to be tested for $g = 2$ is

$$\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = \cdots = \mu_{p1} - \mu_{p2} .$$

Transform the p variates in x to $p - 1$ variates in y as follows (see [1], pp. 110–112 and [12], pp. 239–244):

$$y = \begin{bmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ c_{p-1,1} & \cdots & c_{p-1,p} \end{bmatrix} x$$

such that $\sum_{s=1}^{p} c_{rs} = 0$. The matrix C, subject to the restriction, can be perfectly arbitrary. For example,

$$C = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

subtracts $x_2, \cdots, x_p$ from the first variate resulting in $y_1 = x_1 - x_2$, $y_2 = x_1 - x_3$, $\cdots$, $y_{p-1} = x_1 - x_p$. Or,

$$C = \frac{1}{p}\begin{bmatrix} p-1 & -1 & -1 & \cdots & -1 & -1 \\ -1 & p-1 & -1 & \cdots & -1 & -1 \\ \vdots & & & & & \vdots \\ -1 & -1 & -1 & \cdots & p-1 & -1 \end{bmatrix},$$

which in effect subtracts from each of the p variates their mean $\bar{x} = (1/p) \sum_{i=1}^{p} x_i$ resulting in $y_1 = x_1 - \bar{x}$, $\cdots$, $y_{p-1} = x_{p-1} - \bar{x}$. Using the first transformation above, the vector $y'_{(k)} = (y_{1k}, \cdots, y_{(p-1)k})$ is multivariate normal with mean $\eta'_{(k)} = (\eta_{1k}, \cdots, \eta_{(p-1)k})$, $\eta_{ik} = \mu_{1k} - \mu_{(i+1)k}$, and variance-covariance matrix $C\Sigma C'$, where the prime denotes the transpose of a matrix.

After transforming the p x-variates into the $p - 1$ y-variates for each of the $n = n_1 + n_2$ individuals, the group means in the y's are

$$\bar{y}_{.11}, \bar{y}_{.21}, \cdots, \bar{y}_{.(p-1)1}$$

$$\bar{y}_{.12}, \bar{y}_{.22}, \cdots, \bar{y}_{.(p-1)2},$$

and the pooled sample variance-covariance matrix in the y's, $W = [w_{rs}]$, where

$$w_{rs} = \frac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{n_1} (y_{ir1} - \bar{y}_{.r1})(y_{is1} - \bar{y}_{.s1}) + \sum_{i=1}^{n_2} (y_{ir2} - \bar{y}_{.r2})(y_{is2} - \bar{y}_{.s2})\right\},$$

and $r, s = 1, 2, \cdots, p - 1$. It is easily seen that the null hypothesis in the x's is equivalent to the following hypothesis in the y's:

$$\eta_{j1} = \mu_{11} - \mu_{(j+1)1} = \eta_{j2} = \mu_{12} - \mu_{(j+1)2}, \qquad j = 1, 2, \cdots, (p - 1);$$

i.e., $\eta_{(1)} = \eta_{(2)}$. But this is the general hypothesis of multivariate analysis of the equality of mean vectors for two groups and it is well known that the appropriate statistic to test this hypothesis is T^2. Therefore

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_{(1)} - \bar{y}_{(2)})' W^{-1} (\bar{y}_{(1)} - \bar{y}_{(2)})$$

$$= \frac{n_1 n_2}{n_1 + n_2} \sum_{r}^{p-1} \sum_{s}^{p-1} w^{rs} (\bar{y}_{.r1} - \bar{y}_{.r2})(\bar{y}_{.s1} - \bar{y}_{.s2}),$$

where w^{rs} is element rs in the inverse matrix W^{-1}. This statistic has the T^2 distribution with $n_1 + n_2 - 2$ degrees of freedom.

To test the hypothesis at level α, enter

$$\frac{T^2(n_1 + n_2 - p)}{(n_1 + n_2 - 2)(p - 1)}$$

in the F table with $p - 1$ and $n_1 + n_2 - p$ degrees of freedom. If

$$\frac{T^2(n_1 + n_2 - p)}{(n_1 + n_2 - 2)(p - 1)} > F_\alpha(p - 1, n_1 + n_2 - p)$$

reject the hypothesis; otherwise accept.

The general case for g populations, of which the above is a special case, is given by Heck [9]. The extension is obvious. From the g by $p - 1$ table of group means, one computes the between groups sums of squares and cross products to obtain the elements of the matrix B, say. Thus element rs of this matrix is

$$b_{rs} = \sum_{k=1}^{g} n_k (\bar{y}_{.rk} - \bar{y}_{.r.})(\bar{y}_{.sk} - \bar{y}_{.s.}) = \sum_{k=1}^{g} n_k \bar{y}_{.rk} \bar{y}_{.sk} - n \bar{y}_{.r.} \bar{y}_{.s.}\ ,$$

where $r, s = 1, 2, \cdots, (p - 1)$. For the error matrix W, compute similarly the sums of products, so that

$$w_{rs} = \sum_{k=1}^{g} \sum_{i=1}^{nk} (y_{irk} - \bar{y}_{.rk})(y_{isk} - \bar{y}_{.sk}) = \sum_{k=1}^{g} \sum_{i}^{nk} y_{irk} y_{isk} - \sum_{k=1}^{g} n_k \bar{y}_{.rk} \bar{y}_{.sk}\ .$$

In the above formula, $\bar{y}_{.rk} = n_k^{-1} \sum_{i=1}^{nk} y_{irk}$. The various test statistics proposed are proportional to some function of the product matrix BW^{-1}.

In the literature on multivariate analysis, there have been three approaches to the distribution problem. Wilks [17], starting with the likelihood ratio criterion, derived the test statistic $| I + BW^{-1} |^{-1}$, which is obviously equal to the inverse of the product of the characteristic roots of $(I + BW^{-1})$, I being the identity matrix. Hotelling [10] has proposed the distribution of tr BW^{-1} or of the sum of the characteristic roots of BW^{-1}. Roy [13] has proposed the consideration of the distribution of the maximum characteristic root of BW^{-1}. For a further discussion of these three points of view consult Anderson ([1], pp. 221–224). There are no probability tables available for the first two test statistics although the exact cumulative distribution of the determinantal statistic is given by an infinite series of χ^2's, the first term of which, for any reasonable N, gives an excellent approximation to the whole series and is quite easy to compute ([1], p. 208 and [12], p. 261). Several tables are available giving critical points of the distribution of the maximum characteristic root, the most extensive to date being due to Heck [9].

As an illustration of the T^2 procedure, and also to compare the F tests with the T^2 test, the two smallest groups, B and E, are selected for testing the quality of profile shapes. The mean profiles on the six items are given in Table 3. The analysis of variance performed on these two groups only is displayed in Table 5.

TABLE 5

Analysis of Variance

Source	d.f.	SS	M.S.	F
Tests	5	733.03		
Groups	1	16.25	16.25	
Individuals within Groups	20	111.84	5.59	
Groups x Tests	5	105.56	21.11	$F_3 = 4.84$
Individuals x Tests within Groups	100	435.58	4.36	

To test the hypothesis of no group-test interaction, that is, of equality of the profile shapes, consider F_3, distributed as $F(5\epsilon, 100\epsilon)$. From the pooled sample variance-covariance matrix, an estimate for ϵ is .6727. Therefore, for the approximate test, enter $F_3 = 4.84$ in the F distribution with 3 and 67 degrees of freedom. This test yields

$$.005 > P[F(3, 67) > 4.84] > .001,$$

with

$$P[F(3, 67) > 4.84] \approx .004.$$

The minimum value ϵ can assume is 1/5. Therefore, for the conservative test, enter F_3 in the F distribution with 1 and 20 degrees of freedom. This test yields

$$.05 > P[F(1, 20) > 4.84] > .025,$$

with

$$P[F(1, 20) > 4.84] \approx .04.$$

As previously indicated, to carry out the exact test it is necessary to reduce the dimensionality of the vector from p to $p - 1$. This is accomplished by subtracting x_2, x_3, x_4, x_5, and x_6 from x_1, obtaining a 5-dimensional vector y on each of the 22 individuals. The mean vector for group B is

$$\bar{y}'_B = (4.077,\ .692,\ 3.923,\ 5.692,\ .538)$$

and for group E is

$$\bar{y}'_E = (7.556,\ 4.444,\ 3.889,\ 6.222,\ -.444).$$

The pooled sample variance-covariance matrix in the y's is

$$W = \begin{bmatrix} 5.057 & 2.854 & .382 & .560 & 2.384 \\ \cdot & 7.250 & .906 & 3.994 & 2.196 \\ \cdot & \cdot & 5.090 & 5.696 & 1.155 \\ \cdot & \cdot & \cdot & 12.716 & 3.002 \\ \cdot & \cdot & \cdot & \cdot & 5.272 \end{bmatrix}$$

with $n_1 + n_2 - 2 = 20$ degrees of freedom. Therefore,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_E - \bar{y}_B)' W^{-1} (\bar{y}_E - \bar{y}_B)$$

$$= 27.048.$$

Now enter

$$\frac{T^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2 - p}{p - 1} = \frac{27.048}{20} \cdot \frac{16}{5} = 4.328$$

($p = 6$, the number of x variables) in the F distribution with $p - 1 = 5$ and $n_1 + n_2 - p = 16$ degrees of freedom. This exact test yields

$$.025 > P[F(5, 16) > 4.328] > .01,$$

with

$$P[F(5, 16) > 4.328] \approx .01.$$

By presenting these three tests, no more is implied than that all three tests yield significant results at the .05 level. It is not appropriate to compare the approximate and conservative tests with the exact test based on the distribution of any of the functions of the latent roots previously discussed, all of which reduce to Hotelling's T^2 in the case of two groups. The reason is that there are two exact tests involved; one based on a function of the latent roots, the second based on the distribution of the ratio of linear combinations of χ^2 variables. It should be emphasized that the first two tests given in the example are approximate and conservative for the latter. The conservative test provides a procedure which is more than "rough and ready" and yet saves considerable time since it does not require a matrix inversion nor even the computation of a covariance matrix. This is particularly true when p, the number of variables, is large and the number of samples is greater than two.

The question of electronic computers is another matter. Given the availability of a classical analysis of variance program and the availability of a combined program to carry out the multivariate analysis of variance involving the between samples variance-covariance matrix, the inverse of the error variance-covariance matrix, and the extraction of the maximum latent root of the product of the two matrices, it is very likely that the former would require less machine time. However, the difference is probably of no practical importance and the exact procedure should be used.

A more fundamental question relates to a comparison of the two exact tests involved. Are the multivariate analysis of variance procedures depending upon the distribution of BW^{-1} more powerful against all alternatives than the distribution of the ratio of linear sums of χ^2 variates? It is not clear that this is so, particularly with regard to the analysis of profile shapes where the former procedures must reduce the dimensionality of the random vector.

If one does decide to use the F tests in an analysis, the following series of steps are suggested. After finding the traditional analysis of variance table, first test the appropriate observed F value in the F distribution with full, i.e., unreduced, degrees of freedom. For F_3, for example, this would be F with $(p - 1)(g - 1)$ and $(p - 1)(N - g)$ degrees of freedom. If F_3 is smaller than the α critical point, one can stop here, for the null hypothesis will not be rejected with further manipulation of degrees of freedom. If the observed F is significant, then one proceeds to the conservative test where the degrees of freedom are reduced by a factor equal to $1/(p - 1)$. For F_3, the appropriate F distribution is $F(g - 1, N - g)$. If this test leads to significance at the α level, one can at this point reject the null hypothesis without further testing. However if the conservative test is not significant then it is suggested that the ϵ be estimated from the variance-covariance matrix and the approximate test be carried out.

Number of Individuals Less than the Number of Variables

As indicated in the introduction, in the case of one group, if $(n - 1) < p$, or in the case of g groups, if $(N - g) < p$, it is not possible to apply multivariate procedures. The reason of course is that the error matrix, W, is singular. Such situations are not too uncommon, especially in research in clinical psychology and psychiatry. Clearly the approximate F tests presented are not applicable either since the reduction in degrees of freedom is dependent upon the elements of a singular matrix. However, the conservative test can be applied.

Unequal Variance-Covariance Matrices

Perhaps one of the most important uses of the conservative test is in the situation where one cannot assume the equality of the unknown variance-covariance matrices in the p-variate normal populations being sampled. For this case, there are no exact procedures available. It will be noted that this case, $p = 1$ and $g = 2$, reduces to the Fisher-Behrens problem.

Here, in order for the F statistics to be unbiased, it is necessary to work with equal sample sizes in the groups, i.e., $n_1 = \cdots = n_g = n$. Therefore, $N = \Sigma n_k = gn$. It can again be shown that the respective numerator and denominator quadratic forms entering into F_1, F_2, and F_3 are independent and have the same expectations. Now, however, when an F distribution is used to approximate these F statistics (see [5], theorem 6.1), it turns out that there are different factors reducing the numerator and denominator degrees of freedom, and these in turn differ for the three F statistics. Here again it can be shown that these ϵ's have lower limits which when applied to the appropriate degrees of freedom result in a conservative test for assessing the significance of F_1, F_2, and F_3 by entering these in the F distribution with 1 and $n - 1$ degrees of freedom. It is of interest that the F_2 test, when $p = 1$ and $g = 2$, is a conservative test for the various approximate solutions given to the Fisher-Behrens problem of testing the equality of two means with unequal variances (e.g., [15], p. 295).

REFERENCES

[1] Anderson, T. W. *Introduction to multivariate statistical analysis.* New York: Wiley, 1958.

[2] Block, J., Levine, L., and McNemar, Q. Testing for the existence of psychometric patterns. *J. abnorm. soc. Psychol.*, 1951, **46**, 356–359.

[3] Box, G. E. P. A general distribution theory for a class of likelihood criteria. *Biometrika*, 1949, **36**, 317–346.

[4] Box, G. E. P. Problems in the analysis of growth and wear curves. *Biometrics*, 1950, **6**, 362–389.

[5] Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Ann. math. Statist.*, 1954, **25**, 290–302.

[6] Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. math. Statist.*, 1954, **25**, 484–498.

[7] Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.

[8] Geisser, S. and Greenhouse, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

[9] Heck, D. L. Some uses of the distribution of the largest root in multivariate analysis. Inst. Statist. Univ. North Carolina, Mimeo. Ser. No. 194, 1958.

[10] Hotelling, H. A generalized T test and measure of multivariate dispersion. *Proceedings of the second Berkeley symposium on mathematical statistics and probability.* Berkeley: Univ. Calif. Press, 1951, 23–42.

[11] Kullback, S. An application of information theory to multivariate analysis, II. *Ann. math. Statist.*, 1956, **27**, 122–146.
[12] Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.
[13] Roy, S. N. On a heuristic method of test construction and its use in multivariate analysis. *Ann. math. Statist.*, 1953, **24**, 220–238.
[14] Scheffé, H. A "mixed model" for the analysis of variance. *Ann. math. Statist.*, 1956, **27**, 23–36.
[15] Welch, B. L. Note on Mrs. Aspin's Tables and on certain approximations to the tabled functions. *Biometrika*, 1949, **36**, 293–296.
[16] Wilk, M. B. and Kempthorne, O. Fixed, mixed, and random models. *J. Amer. statist. Ass.*, 1955, **50**, 1144–1167.
[17] Wilks, S. S. Certain generalizations in the analysis of variance. *Biometrika*, 1932, **24**, 471–494.

Manuscript received 8/21/58

Revised manuscript received 12/1/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

DOMAIN SAMPLING FORMULATION OF CLUSTER AND FACTOR ANALYSIS

ROBERT C. TRYON

UNIVERSITY OF CALIFORNIA, BERKELEY

Domain sampling principles permit formulation of a general method of multidimensional analysis. Cluster and factor analysis methods are special cases stemming from decisions made at different stages of the general method, especially in defining an independent dimension. Key cluster analyses define a dimension as a selection of s variables drawn from the full n set. Centroid, principal axes, and maximum likelihood analyses define it by the n variables (raw or residual, weighted or unweighted); bifactor and second-order analysis, by both types of selection; square root analysis, by one variable. Key cluster methods can be designed to test hypotheses.

The definition of a variable as the sum of a sample set of scored responses (e.g., to test items) selected to be representative of a defined domain of behavior is a basic principle of psychometrics. This standard practice may be expressed in a simple algebraic fashion which leads to an integration of the plethora of formulations of the reliability coefficient [39]. When a test is included among n variables, domain sampling algebra also provides a definitive solution of its communality [40]. These principles have been shown to implement the broad logic of multidimensional analysis by the psychometric procedures called cluster analysis [42]. The most generally applicable computational variant of cluster analysis, the CC method, has also recently been published [41].

Completing this group of papers on domain sampling formulations, this article has as its purpose, first, to state the general case of multidimensional analysis, and, second, to develop from it important special cases that are variant methods of cluster analysis. Some of these special forms have, however, been otherwise known over the last half century as factor analysis methods, their main originators being Spearman [29, 30], Thomson [31], Burt [2, 3], Kelley [19, 20], Hotelling [14, 15], Thurstone [32, 34], Holzinger [13], and Lawley [21]. The factor methods of Spearman, Kelley, Thurstone, and Holzinger are conceived as issuing from the basic factor theorem. The assumptions are that a test score results from underlying, uncorrelated and additive true (general and multiple), specific, and error factors. These restrictive assumptions of factor theory are difficult to justify on substantive biological and psychological grounds [36]. This paper shows that when the factor methods are recast as variants of cluster analysis such assumptions about the components of a test score are unnecessary restrictions.

TABLE 1

DECISIONS THAT DISTINGUISH VARIOUS METHODS OF MULTIDIMENSIONAL ANALYSIS

REGIONS OF DECISION	KEY CLUSTER				PIVOT VARIABLE	GEN'L AND KEY	GENERAL CLUSTER		
	1	2	3	4	5	6	7	8	9
	Resid Cluster		Precluster		Factor Analysis Variants				
	Tot Com (TC)	Cum Com (CC)	Preclus Cum Com (PCC)[a]	Ration Cum Com (RCC)	Square Root (PV)	Bi-factor	Cent-roid (Unwgt)	Prin-cipal Axes (Wgt)	Maxi-mum Like-lihood
PRELIM DECISIONS									
a Reflection	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Factor					*No*	*No*	*Yes*	*No*	*No*
b Precluster	No	No	Empiric	Ration	Yes	No	No	No	No
Factor					*No*	*Yes*	*No*	*No*	*No*
DIAGONAL ENTRIES									
c Initial Communalities	Simul Quad	Approx	Approx[b]	Approx[b]	Approx	Approx	Approx	Approx	No
Factor					*Unities*	*No*	*High r*	*Unities*	*No*
DIMENSIONALITY									
d Defining Variables of C_x	s_x	s_x	s_x	s_x	3 on 1	n and s_x	n	n(wgt)	n(wgt)
Factor					*1*	*k and* s_x	*n*	*n(wgt)*	*n(wgt)*
e Partial Communalities	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Iter	Iter
Factor					*Sim* Σ[c]	*Appr B*	*Sim* Σ	*Iter*	*Iter*
f Terminating Criteria[d]	T	T	T	T	T	T	T	T	Resid r
Factor					*Salient*	*k + 1*	*Resid r*	*Resid r*	*Resid r*
g Reiterate Factoring	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Prog k
Factor					*No*	*No*	*No*	*No*	*Prog k*
STRUCTURE									
h Oblique Analysis	From d	From d	From b,d	From b,d	From d	From d	Rotate	Rotate	Rotate
Factor					*No*	*No*	*Rotate*	*No*	*No*
SCORING									
i Oblique Dimension Scores	Σ	Σ	Σ	Σ	Σ	Σ	Regress	Regr	Regr
Factor					Σ	*Regress*	*Regr*[e]	*Regr*	*Regr*

[a] For factor analysis methods (e.g.,"multiple group"), see text

[b] In abridged form of the method, no initial communalities are necessary

[c] Specified for the special case of $s_x=1$ and communalities unity

[d] Sampling criteria may be used in place of or in addition to the T criterion

[e] Regression in theory; cluster score by Σ in practice

The broad plan of this treatment is represented schematically in Table 1. The general case of multidimensional analysis consists of progressive stages, or regions, in each of which decisions by the analyst are required. These regions are summarily described in the lettered rows a to i, extreme left column of Table 1. The numbered columns 1 to 9 are the special cases, each defined by the particular pattern of decisions listed down its column. The first four cases form the *Key Cluster* methods, columns 1 to 4. The remaining groups are other cluster analysis variants known also as factor analysis methods (or schools): their descriptive names are given below the column numbers. For example, column 7 defines the Thurstone centroid–simple structure method. In this group of methods, note that in each region of decision (or row) there are two entries. The first entry is a decision resting on the conception of the method as a variant of cluster analysis. The second, in italics and labelled at the left as "Factor," is the decision made in orthodox factor formulation. For example, down column 7 Thurstone's decisions are represented by the pattern of second, italicized, entries.

The general case of multidimensional analysis is given in the next section of the paper. Each region of analysis is taken up in order, the principles involved being illustrated by referring to some of the types of decisions listed along its row. In later sections the special cases will be taken up successively, that is, for each column of Table 1 the nature and rationale of its pattern of final decisions will be given, first, conceiving the method as a variant of cluster analysis (first cell entries), and, second, as an orthodox factor formulation (second cell entries).

General Case of Multidimensional Analysis

The basic data of the analysis are the intercorrelations between scores on variables $X_1, X_2, \cdots, X_v, \cdots, X_n$. The over-all objective is *to determine and measure the smallest number, k, of dimensions that reproduce the intercorrelations*, entered as side elements in a correlation matrix. The successive stages of the analysis, rows a to i of Table 1, achieve this objective. For convenience, these stages are grouped under five subordinate objectives, lettered A to E below.

A. *Preliminary Decisions* (Table 1, rows a, b)

Reflection of Variables (row a). A main desideratum in deciding on whether to reflect variables is the method of computing partial communalities (squared factor loadings), row e. In the special cases of columns 1 to 7, where the simple summation (Sim Σ) formula (to be developed later) is used, reflection of variables, denoted by "Yes" in row a, is required.

Preclustering Variables (row b). Preclustering variables before dimensional analysis permits certain abbreviated or rational variants of multidimensional analysis (columns 3, 4, 5). In the other variants test clusters are located during the dimensional analysis (1, 2, 6) or by rotation of dimensions (7, 8, 9).

B. *Determining Initial Communalities* (Table 1, row *c*)

The diagonal entries of the matrix are "self-correlations." Communalities are selected as diagonal elements because their use yields the smallest number, k, of necessary and sufficient dimensions, or uncorrelated cluster domain scores, that will reproduce the intercorrelations. If one used the reliability coefficient, $r_{tt'}$, of each variable as its diagonal element, more dimensions than k would be required, and if one set each diagonal value at unity, even more dimensions would be necessary.

As a correlation coefficient, the communality h_v^2 of a variable v is defined as the correlation between the observed variable X_v and a hypothetical construct variable $X_{v'}$ measuring a *different* behavior property than X_v but having correlations across the $n - 1$ other variables equal to those of X_v ([40], formula 1) i.e.,

$$h_v^2 = r_{vv'} \,. \tag{1}$$

As a variance, the fundamental definition of h_v^2 is the proportion of total variance of the observed scores of v that is predictable from the construct variable-domain score, C_v, defined as

$$C_v = z_v + z_{v'} + \cdots + z_{v\infty} \,, \tag{2}$$

in which the observed z_v-scores are defined as one sample variable drawn from an infinite set of construct variables, all members of which measure different behavior properties, but whose correlations with the $n - 1$ variables are proportional to those of v. From this construction it follows that ([40], formula 5)

$$h_v^2 = r_{vC_v}^2 \,. \tag{3}$$

Proportionality of the correlations of v with those of another variable i means specifically

$$r_{vj}/r_{ij} = \text{a constant} \qquad (j = 1, \cdots, n; v \neq i \neq j). \tag{4}$$

Taking the following as an index of proportionality ([41], formula 6; [43]),

$$P_{vi}^2 = \frac{(\Sigma r_{vj} r_{ij})^2}{\Sigma r_{vj}^2 \Sigma r_{ij}^2} \qquad (j = 1, \cdots, n; v \neq i \neq j), \tag{5}$$

then under condition (4), P_{vi}^2 is unity. The definition of the variable-domain C_v in (2) is not restricted to a domain of variables with equal correlations but merely to those with proportional r's, as defined in (4).

Communalities are in practice not computed by their defining formulas (1) and (3) because the requisite construct variables are not available. In

Table 1, row *c*, note that in TC analysis (to be described later) a solution is attempted by a quadratic equation. But in remaining methods approximations are taken; after reiteration of the factoring procedure, row *g*, final converged values are achieved.

C. *Determining Dimensionality* (*Factoring*) (Table 1, rows *d*, *e*, *f*, *g*)

The object is to determine the value of k, the number of uncorrelated composite variables or independent dimensions, $C_1, C_2, \cdots, C_x, \cdots, C_k$ that could reproduce the correlation matrix, including diagonal communalities.

(a) The communality, h_v^2, of any variable v may be partitioned into k partial communalities (squared factor loadings) or h_{xv}^2 values as follows:

$$h_v^2 = h_{1v}^2 + \cdots + h_{xv}^2 + \cdots + h_{kv}^2 . \tag{6}$$

(b) The correlation coefficient of v with each other variable i is reproduced, i.e., the observed r_{vi} equals

$$r'_{vi} = h_{1v}h_{1i} + \cdots + h_{xv}h_{xi} + \cdots + h_{kv}h_{ki} , \tag{7}$$

or, said another way, each of its residual correlations after removing the variance from dimensions 1 to k is

$${}_{1\ldots k}r_{vi} = r_{vi} - r'_{vi} = 0. \tag{8}$$

Definition of an Independent Dimension, C_x (*An Orthogonal Factor*) (Table 1, row *d*). The score C_x is a composite, defined as the following independent cluster domain score:

$$C_x = (w_a)_r C_a + (w_b)_r C_b + \cdots + (w_{s_x})_r C_{s_x} . \tag{9}$$

The *defining variables* of the dimension are selected observed variables a, b, $\cdots$, s_x, taken from all n variables. The C's are their variable-domain scores, defined by (2). The prescript r means that scores on preceding dimensions $C_1, \cdots, C_{x-1}$ are held constant, thus establishing the independence of C_x. The w's are weights. The same variable may appear in different dimensions, though, of course, as different residual scores.

This definition is more general than that written originally by Pearson [25], who initiated multidimensional analysis. It encompasses as special cases the varieties of cluster and factor analysis given in Table 1. In row *d*, note that in key cluster analysis (TC, CC, PCC, RCC), each dimension is defined by a cluster of s_x variables usually less than n. In centroid, principal axes, and maximum likelihood factor analysis, the defining variables are indiscriminately a general cluster of all n variables. These latter methods differ from each other in the values of the weights attached to the different component variables in (9). In bifactor analysis, dimension C_1 is a general cluster of all n variables, but later dimensions are key clusters. In pivot variable ("square root") analysis, each dimension is defined focally by one variable only.

Recall that each C-variable is a variable-domain defined in (2) as a composite of z-scores of variables measuring different behaviors but showing proportional correlations, hence the use of communalities in the diagonals. Were the z-scores in (2) defined as different test samples of the *same* behavior domain, each C-variable would be the construct composite called the "true" or X_{t_∞}-score of the variable [39] and reliability coefficients would be diagonal entries of the matrix. Were each C-variable defined as the single z-score of the defining variable, then the diagonals would be unities. In multidimensional analysis, the definition leading to communalities in the diagonal is chosen for the reason given under objective B.

The simplest weighting of the C-variables is to set all w's to unity in (9), as in TC, CC, PCC, RCC, centroid, diagonal, and bifactor analysis. This simple summation, as Burt calls it [3, 4], intrinsically weights each defining variable of the dimension C_x by its proportional contribution to the variance of C_x, that is, by the sum of its communality and its correlations with the other defining variables. Another choice, characteristic of the principal axes and maximum likelihood methods, is the computationally arduous least squares solution of the sets of weights which yield the maximal sum of their h_x^2 values, that is, of their partial communalities.

Partial Communalities (Squared Factor Loadings) (Table 1, row *e*). The portion of the variance of a variable v predictable from a dimension C_x is the square of its correlation with C_x, called its partial communality, $h_{x_v}^2$, defined in (6). In the unweighted case, from the correlation of sums in the limit,

$$(10) \qquad h_{x_v}^2 = r_{vC_x}^2 = \frac{(\Sigma_r r_{vi})^2}{\Sigma_r h_i^2 + 2\Sigma_r r_{ij}} \qquad (i, j = a, \cdots, s_x\,; i < j).$$

The numerator is simply the square of the sum of the residual correlations of v with the defining variables of C_x. The denominator is simply the sum over the total submatrix of residual correlations of these defining variables, including the diagonal residual communalities. When $i = v$, then ${}_rh_i^2$ is included in the numerator.

A residual correlation, from (7) and (8), is

$$(11) \qquad {}_rr_{vi} = {}_{1\ldots(x-1)}r_{vi} = r_{vi} - (h_{1_v}h_{1_i} + \cdots + h_{(x-1)_v}h_{(x-1)_i}),$$

and similarly for the ${}_rr_{ij}$ terms. A residual communality is, from (6),

$$(12) \qquad {}_rh_i^2 = {}_{1\ldots(x-1)}h_i^2 = h_i^2 - (h_{1_i}^2 + \cdots + h_{(x-1)_i}^2).$$

If one has chosen the defining variables of C_x before the dimension analysis as in PCC and RCC analysis, it is *not* necessary to work out the individual residual terms of (11) and (12). Only sums from the raw matrix plus diagonals and sums of partial communalities on prior dimensions are necessary. Thus, from (11) the numerator of (10) is

$$\Sigma_r r_{vi} = \Sigma r_{vi} - (h_{1v} \Sigma h_{1i} + \cdots + h_{(x-1)v} \Sigma h_{(x-1)i}) \qquad (i = a, \cdots, s_x). \tag{13}$$

When $i = v$, h_v^2 is included in Σr_{vi}. The denominator of (10),

$$\Sigma_r h_i^2 + 2\Sigma_r r_{ij}^2 = \Sigma r_{ij} - [(\Sigma h_{1i})^2 + \cdots + (\Sigma h_{(x-1)i})^2] \qquad (i, j = a, \cdots, s_x). \tag{14}$$

Σr_{ij} is the sum over the raw matrix of s_x variables including diagonals.

This simple general formula (10), called "Sim Σ" in Table 1, row *e*, is used by all cluster and factor analysis methods, excepting principal axes and maximum likelihood (to be considered later). Recall that for different methods, it differs only in the value of s_x, e.g., in centroid analysis, $s_x = n$.

Terminating Criteria (Salient Dimension Analysis) (Table 1, row *f*). As a simple rational standard for terminating factoring, the writer proposes the *communality exhaustion criterion*. To end factoring by this criterion, one estimates the communalities of all the variables at the beginning of the analysis. Factoring then proceeds up to the dimension C_k at which the communalities of the n variables are exhausted.

Recall that the communality h_v^2 is basically defined in (3), quite independent of the dimension analysis, as the variance of variable v predictable from its variable-domain C_v. This magnitude is estimated before factoring is undertaken by computing formulas given later in the paper, (31) or (32). After factoring is under way, the variance of v predictable from dimensions $C_1, C_2, \cdots, C_x$ is represented by h_{vx}^2, this magnitude being the sum of partial communalities of v up to and including h_{xv}^2, as shown in (6).

Writing the ratio of these two variances, i.e.,

$$F_{xv} = \frac{h_{vx}^2}{h_v^2}, \tag{15}$$

factoring may be terminated on that dimension C_k at which the numerator of (15) approaches the denominator, i.e., when

$$F_{kv} = \frac{h_{vk}^2}{h_v^2} = 1.000. \tag{16}$$

To evaluate F for each variable on each dimension as factoring proceeds would, however, be a complex procedure. Furthermore, the magnitude of F for a given variable is subject to substantial error, both in the initial approximation to the denominator term and in the initial dimensional estimate of the numerator. Less subject to these errors is the approximate sum (or average) of the F-values over all n variables, namely,

$$T_x = \frac{\Sigma h^2_{vx}}{\Sigma h^2_v}. \tag{17}$$

Using T_x as a criterion, one stops factoring on the dimension C_k for which T_k by (17) first equals or exceeds unity, i.e., where

$$T_k = \frac{\Sigma h^2_{vk}}{\Sigma h^2_v} \geqq 1.000. \tag{18}$$

In practice, the T-criterion in (18) appears to yield the most *salient* dimensions [41]. Consider bias in T: initial estimates of communalities are usually biased downward, suggesting that a salient terminal dimension might be rejected by T. This is unlikely because such a dimension would be the one for which T first *exceeds* 1.000. The effect of an upward bias would be to accept nonsalient dimensions. To minimize such an effect, the analyst may set the criterion a little under unity, say, at .975.

The T-criterion may reject later dimensions that would be accepted on sampling grounds. Many analysts may consider such rejection to be an unimportant loss, for such dimensions contribute minor general variance and are usually difficult to interpret.

Significant Dimension Analysis. On the basis of *sampling criteria,* one may wish, however, to accept all significant (as distinguished from salient) dimensions. The orthodox F-test procedures applied to the communality exhaustion indices represented by (15) and (17) would seem appropriate, but their sampling characteristics are not yet known. There remain the various significance tests applied by factorists to the distribution of residual correlations (11) or to the distribution of the square roots of the partial communalities (i.e., of the factor loadings) of a given dimension (10). The tests developed by Saunders are of special interest because he proposes both types ([6], formulas 44 and 46, p. 300ff).

Reiteration of the Factoring to Converged Communalities (Table 1, row g). After the first factoring is complete, the sum of partial communalities by (6) may not yield the correct value of the communality of each variable, as in artificial or population matrices [40]. In those methods that start with approximations, reiteration of the dimensionality analysis on the k dimensions is required until convergence is secured, as shown in row g, Table 1.

The decision as to which decimal place will define convergence may be made on arbitrary grounds of salience, say, the third place. On sampling grounds, however, one may terminate convergence when for every variable the difference between two successive iterated values of its communality by (6) becomes less than, say, a third of the standard error of the last iterated value of its communality. Treating h_v as a multiple correlation, as in (38), later, the approximate error is

$$\sigma_{h_v} \doteq (1 - h_v^2)/\sqrt{N - (k + 1)} \doteq (1 - h_v^2)/\sqrt{N - 2}. \tag{19}$$

The magnitude of k, at least 1, is usually trivial relative to N, hence leading to the final approximation shown.

D. *Determining the Structure of the Interrelationships* (Table 1, row h)

Having determined the dimensionality of the intercorrelations, one may relax the condition of independence and select or derive the k dimensions that may be oblique to each other and be better defined by the observed variables. Key cluster analysis routinely locates those groups of variables which delineate the k most nearly independent oblique dimensions. Corre- sponding to the independent dimensions C_1 , C_2 , $\cdots$, C_k there are the matched set of oblique dimensions, respectively, C_I , C_{II} , $\cdots$, C_K . Thus for independent dimension C_x given in (9) there is an oblique dimension C_y defined in simple summation form by the domain score

$$C_y = C_a + C_b + \cdots + C_{sx} . \tag{20}$$

Scores on the C-variables that form this composite have the same definition as in (9), but they are *not* residual scores as in (9). Geometrically, dimension C_y is an oblique subcentroid in k-space.

If the analyst wishes to check on the clusterings indicated by the dimensional analysis with an eye, perhaps, to a possible reclustering of the variables including those that had remained unclustered, he may employ a geometric model (for an illustration, see [41], Fig. 1). This model takes the k dimensions as independent axes, and each variable as a point on them. The coordinate of each variable on any axis C_x is its correlation, r_{vC_x} (unaugmented factor loading), which by (10) is

$$r_{vC_x} = h_{xv} . \tag{21}$$

The resulting interior model of variable-points is perceptually not as descriptive of the interrelationships among them as the surface model, given by plotting each variable-domain C_v by its augmented correlation, this being, from the correlation of sums in the limit,

$$r_{C_vC_x} = h_{xv}/h_v . \tag{22}$$

The surface model, in which all variable-domains are points at distance 1.00 from the origin, has the perceptual merit of revealing directly as a surface separation of points the relationship between each variable-domain C_v and any one of the other variable-domains C_i . This important "common factor correlation" can, however, be computed directly from the matrix and the communalities, i.e., from the correlation of sums in the limit,

$$r_{C_vC_i} = r_{vi}/h_vh_i . \tag{23}$$

Simple Cluster Structure (*Rotated Primary Factors*). The finally selected cluster domains of type (20) are the most nearly independent k dimensions evident in the data. The degree of their interdependence is given by their intercorrelations. For any two such generally oblique cluster domains, C_{yi} , C_{yj} , their correlation is given, from the correlation of sums in the limit, as

$$r_{C_{yi}C_{yj}} = \frac{\Sigma r_{ij}}{\sqrt{\Sigma r_{ii}}\ \sqrt{\Sigma r_{jj}}}, \tag{24}$$

where Σr_{ii} is the sum of r's in the submatrix of the s_i variables of C_{yi} including diagonal communalities, Σr_{jj} is the same for the submatrix of s_j variables defining C_{yj} , and Σr_{ij} is the sum of the $s_i s_j$ coefficients in their cross correlation submatrix. Recall that, as stated under (9), a given variable may appear in more than one oblique dimension, a situation which would, of course, increase obliqueness.

As an aid in interpreting an oblique dimension C_{yi} , one may compute the correlation of each known observed sample variable v with the dimension (rotated factor loading). By the correlation of sums in the limit it is

$$r_{vC_{yi}} = \frac{\Sigma r_{vi}}{\sqrt{\Sigma r_{ii}}}. \tag{25}$$

But of more interest theoretically is the correlation of C_{yi} with each *kind* of general variation of which v is taken as a test sample, namely, its variable-domain C_v in (2). This correlation (augmented rotated factor loading) is simply, from the correlation of sums in the limit,

$$r_{C_vC_{yi}} = r_{vC_{yi}}/h_v . \tag{26}$$

Difficult problems arise in simple cluster structure analysis in those methods in which the defining variables of the independent C_x dimensions are total clusters of all n variables. As shown in Table 1, row h, columns 7, 8, 9, these dimensions must be rotated [see 9] to meaningful defining orthogonal or oblique clusters. Orthodox factor analysts following Thurstone [34] propose graphical rotation—a cumbersome, subjective procedure, admittedly an art [6]. Recent attempts have been made to achieve rotation by objective analytic methods [5, 18, 24, 26, 28]. Rotation, an unnecessary burden, is not required when the dimensions are defined by key clusters.

E. *Scores on Oblique Dimensions* (Table 1, row *i*)

Ideally, the best estimate of an individual's sample score C_{ys} on any dimension C_y is the regression of C_y on the n variables, i.e.,

$$\bar{C}_{ys} = \Sigma\beta_{C_{yi}}z_i \qquad (i = 1, \cdots, n). \tag{27}$$

Such an estimate is so arduous to compute that in most factor analyses the important job of measuring persons on the reduced dimensions is rarely tackled, and a main benefit of the analysis is lost.

When, however, a dimension is defined by a key cluster, a good estimate may be secured from the cluster score, namely, the simple sum of the z scores of the defining variables of C_v, i.e.,

$$C_{vs} = z_a + \cdots + z_s . \tag{28}$$

In this composite each defining variable takes an intrinsic weight proportional to the sum (or mean) of its correlations with the remaining defining variables. In Table 1, row i, the simple summation, labelled "Σ" may be used in all but the general cluster methods (columns 7, 8, 9).

The *cluster domain validity* of the observed cluster score (28) is its correlation with the full domain score, C_v in (20). By the sums formula in the limit it is

$$r_{C_{vs}C_v} = \sqrt{\frac{\Sigma h_i^2 + \Sigma r_{ij}}{s + \Sigma r_{ij}}} \qquad (i, j = a, \cdots, s; i \neq j). \tag{29}$$

Note that the numerator in (29) is simply the sum over the submatrix of C_v including diagonal communalities, and the denominator is the same except with unities in the diagonals.

The relationships between cluster scores that fallibly measure the final k oblique dimensions are given by their intercorrelations. Between any two such scores, C_{vis}, C_{vjs}, by the sums formula this correlation is

$$r_{C_{vis}C_{vjs}} = \text{same as (24) but with the unities in the diagonals.} \tag{30}$$

Special Cases of Multidimensional Analysis

Key Cluster Analysis: Total Communality (TC) and Cummulative Communality (CC) (Table 1, columns 1, 2)

The TC and CC methods directly apply the general formulations outlined above. As shown in Table 1, columns 1 and 2, the correlation matrix is initially made positive (row a) by conventional reflection methods (see [10], Table 16.13) in order to guarantee that variables chosen to define a given dimension will show correlations of positive proportionality in (4).

An electronic computer is required in TC analysis to solve for the communalities by a simultaneous quadratic formula (row c). Difficulty has, however, been experienced in achieving a solution in empirical matrices [16, 17]. The CC method has therefore been developed [41] to meet the possible failure of solutions by the quadratic formula. CC analysis starts with approximations to communalities, and dimensionality analysis is reiterated until convergence is secured. CC analysis is thus a procedure alternative to TC, being a method that provides a solution in all matrices, and one that may be programmed either for electronic or desk calculator computation.

Solution of the communalities in TC analysis is based on the fact that the communality of any variable v is the squared multiple correlation between v and the remaining $n - 1$ variable-domains ([40], formula 11),

$$h_v^2 = R^2_{v.C_1C_2 \ldots C_i \ldots C_n} \qquad (i \neq v). \tag{31}$$

Solving for the communalities of all variables requires a simultaneous solution by reiteration of the n quadratics of type (31).

Various initial approximations to communalities required in the CC method are available [40, 45, 46]. On domain sampling grounds, the preferred estimate of h_v^2 is "Approximation B" ([40], formulas 29, 30), the Spearman formula, computed from a cluster of reference variables whose correlations are most nearly proportional to those of v. Here,

$$h_v^2 = \Sigma r_{vi} r_{vj} / \Sigma r_{ij} \qquad (i, j \neq v; i < j), \tag{32}$$

where both i and j are the three variables showing the highest P^2_{vi} and P^2_{vj} values, respectively, by (5).

In both methods the defining variables of each dimension C_x anchor on a selected pivot variable, v_1, that appears to show relatively high *and* low correlations with other variables. Such a variable would usually center the cluster obliquely to clusters defining other dimensions. To locate the pivot variable a measure of "pivotness" of each of the n variables is first computed, namely, the variance of its squared r's. The pivot variable v_1 then is that variable whose

$$\text{var}\,({}_r r^2_{v_1 i}) \text{ is the maximum} \quad (i = 1, \cdots, n; i \neq v). \tag{33}$$

A quicker and probably less sensitive means of selecting the pivot variable is to choose the one with the highest residual communality as given in (12). For desk calculator work, selecting v_1 from a correlation distribution table is satisfactory, though it entails some subjective elements ([41], Table 2).

Around the pivot variable one collects the remaining defining variables of C_x. These are the variables with highest indices of proportionality with v_1. Three such variables at a minimum are selected. If these variables are called, in order of magnitude of $P^2_{v_1 i}$, i_1, i_2, i_3 then any additional variable i may also be included in the cluster if its $P^2_{v_1 i}$ value is equal to or above .81 and also is within twice $P^2_{v_1 i_1} - P^2_{v_1 i_3}$, that is, if its

$$P^2_{v_1 i} \gtreqless (2P^2_{v_1 i_3} - P^2_{v_1 i_1}) \gtreqless .81. \tag{34}$$

Partial communalities, row e, are computed by the general formula (10). In TC analysis the factoring process is terminated by the T-criterion at the end of one factoring procedure only. But in CC analysis, since approximations initiate the analysis, the first factoring process is terminated by the T-criterion, then new values of the communalities are computed from (6), and the factoring process is reiterated until the communalities converge (rows f, g). In both TC and CC analysis, the dimensional analysis locates the

k oblique dimensions, row h, and scores of individuals on the dimensions, row i, follow the formulations as given earlier under the general method.

Key Cluster Analysis: Preclustered Cummulative Communality (PCC) and Rational Cummulative Communality (RCC) (Table 1, columns 3, 4)

Recall that in TC and CC analyses the cluster of variables selected to define a dimension C_x is chosen *during* the dimension analysis. One may, however, choose them *prior* to factoring, empirically in PCC, rationally in RCC analysis. Preselection of clusters makes multidimensional analysis a quick desk calculator operation because the complete residual correlation matrices essential to CC analysis are not required. PCC and RCC analyses are procedurally identical after the analyst has clustered the variables (see Table 1, columns 3 and 4).

In PCC analysis (for a recent illustration see [38]), one *empirically* groups the n variables into k' clusters, $C_I, \cdots, C_y, \cdots, C_{K'}$. Each cluster is made to be as "tight" as possible, i.e., is composed of variables whose correlations are maximally proportional by (5). Some variables may remain unclustered, but their number is kept as small as possible. As an aid in selecting the groupings one may use a correlation distribution table.

But in RCC analysis the rational groupings stem from the analyst's *theory* from which he generated the n variables under study. An investigator commonly conceives the n variables to sample different behavior domains or properties of the individuals, such as the facets of Guttman [11]. The s variables that fall in each such theoretical subgroup are a rational cluster. The n variables will usually be organized in k' such clusters, though a few may remain as isolates.

As in CC analysis, one starts both PCC and RCC analyses by computing approximations to the communalities, preferably by formula (32). Thereafter the work is procedurally identical to the CC analysis excepting that only mean residuals are necessary. The mean residual correlation of a variable v with any cluster C_y is, from (13),

$$(35) \qquad {}_r\bar{r}_{vi} = (1/s_y)\Sigma_r r_{vi} \qquad (i = a, \cdots, s_y).$$

The mean residual communality of the variables that compose C_y is, from (12),

$$(36) \qquad \overline{{}_rh_i^2} = (1/s_y)\Sigma_r h_i^2 \,.$$

The quickest means of choosing the *pivotal defining cluster* of any dimension C_x is to select the one with largest mean residual communalities as given in (36). The partial communalities are then computed by the simple summation formula (10). As shown in Table 1, the remainder of PCC and RCC analyses is the same as in CC. The final k dimensions are thus defined by a selection from the k' original clusters.

A complication may arise if one should run out of clusters before dimensionality has been completely determined. If one wishes precision on dimensionality he would compute the $n \times n$ residual correlation matrix and would perform the CC procedures on it and on any later residual matrices that are necessary. PCC and RCC analyses can be made as precise as CC analysis. To do so the analyst forms new estimates of the communalities by (6) and then, as in CC analysis, reiterates the factoring procedures until communalities converge.

Abridged PCC and RCC (*Multiple Group Factor Analysis or "Poor Man's Cluster Analysis"*). The analyst need only spend a few hours of work on a correlation matrix if he is satisfied with an approximate multidimensional analysis. After reflecting the variables, preclustering them, and approximating their communalities as in PCC analysis proper he can compute the correlations between the resulting k' cluster domains by (24). He may, if he wishes, even *skip* the step of approximating communalities, leave the diagonal vacant, and use mean r's instead of Σr's in (24).

The result is a $k' \times k'$ matrix with diagonal elements of unity. The correlation between each variable v and a given domain C_y is then estimated by (25); if communalities are not used, he would use mean r's instead of Σr values in (25). These calculations result in an $n \times k'$ matrix. From a study of these data he may make a reasonable estimate of the dimensionality, interpret the oblique dimensions, and compute cluster scores. If he wishes a more accurate estimate of the dimensionality and of the most oblique k clusters, he can factor the $k' \times k'$ matrix by the diagonal method of factoring (see "Pivot variable analysis" below). By this means he quickly locates the k most nearly independent cluster dimensions.

Orthodox PCC Analysis (*Group, Grouping, and Multiple Group Methods of Factoring*). In 1939, the writer published dimensional analysis by the PCC method in approximately the form presented here ([37], Sec. 7); the quick abridged form was also given (Sec. 5, and Analyses 14, 15a for centroids). Five years later Holzinger [12] and then Thurstone [33, 35] presented the abridged form. Thurstone labels it the multiple group method of factoring. At the end of abridged analysis he rotates the k oblique dimensions to orthogonal positions in order to compute residual correlations, and to see if further dimensions might be necessary. In his book, Thurstone [34] added the group and the grouping methods (see also [6], ch. 11). They differ from the 1939 and current PCC methods in the procedure of grouping variables in clusters on grounds of absolute *magnitudes* of correlations rather than of proportionality of correlations, our P^2 criterion.

Pivot Variable (PV) Cluster Analysis (*Diagonal or Square Root Factor Analysis*) (Table 1, column 5)

In PV analysis (see Table 1, column 5), each variable i among the n

variables may be selected as the central sample variable of an oblique cluster, C_{v_i}. Defined generally in (20), C_{v_i} here consists of three variables only. The other two variables are those which, after the matrix is reflexed, yield the highest P^2 values with i. On domain sampling principles one may thus conceptualize $k' = n$ preclustered domains, $C_{v_I}, \cdots, C_{v_i}, \cdots, C_{v_n}$. After determining approximations to the communality of each variable from its reference variables by (32), one computes the correlations between the n domains by (24). Then a dimensional analysis of the $n \times n$ matrix of $r_{C_{v_i}C_{v_i}}$ coefficients, with diagonal elements of unity, is performed. The cluster with highest column sum defines the first dimension; locating the pivot cluster of later dimensions requires only computing residual communalities and selecting the one with highest value. With an electronic computer, however, one can instead compute residual matrices and more sensitively select the pivot variable by (33). If the defining pivot cluster of any dimension C_x is denoted C_{v_p}, then the augmented partial communality of an oblique cluster is a special case of (10), i.e.,

$$h^2_{xC_{v_i}} = {}_r r^2_{C_{v_i}C_{v_p}} / (1 - {}_r h^2_{C_{v_p}}). \tag{37}$$

The numerator of (37) calls only for the residual interdomain correlations of the selected pivot cluster with each of the remaining $n - 1$ clusters; only a simple $n \times 1$ matrix of such residual correlations is necessary. Factoring is terminated when the sum of the augmented partial communalities of all variables over all dimensions, that is, the numerator of the T-criterion, first becomes equal to or greater than the denominator term, n.

Having now located k pivot clusters by the dimensional analysis, the analyst assigns each of the remaining $n - k$ variables to a final set of k oblique clusters. Each may be assigned to that pivot cluster which defines the dimension on which the variable in question has its highest partial communality. Tighter clusters may, however, be grouped by criterion (34). An illustration of PV diagonal factoring procedures applied to an interdomain matrix is given elsewhere ([38], Appendix D). The potentialities of PV analysis should be explored. It is a rapid means of estimating k, the dimensionality (rank) of the matrix, and hence it could precede maximum likelihood analysis where foreknowledge of the approximate value of k is desirable.

PV analysis would give precise results if one started with correct values of the communalities. But it may be employed to find such values, as follows. After the first factoring to determine the k most nearly independent pivotal clusters, these clusters may be used as a constant reference set of predictors to compute the communality of each variable, v. Elsewhere it has been shown ([40], formula 44) that h_v^2 is the squared multiple correlation between v and a set of k oblique cluster domains, i.e.,

$$h_v^2 = R^2_{v.C_I C_{II} \cdots C_K}. \tag{38}$$

From knowledge of the k oblique clusters by initial PV analysis, and with initial trial values of the constants required in (38) computed from (32), (24), and (25), simultaneous solutions of all n communalities, reiterated to convergence, can be programmed electronically. Such a program may be integrated with a periodic replication of PV analysis in order to discover whether the increased precision of the communalities produces changes in dimensionality.

Orthodox PV Analysis. One of the oldest factoring methods, diagonal factoring, used in PV analysis has recently been relabelled square root factor analysis [44]. In orthodox practice the analyst pivots a dimension on a *single* variable by arbitrarily inserting unities in the diagonal and factoring the unreflected r matrix. This practice prevents one from determining the dimensionality of the matrix, rests the partial communalities rather unstably on the coefficients of the pivot variable alone, and leaves unclear the oblique cluster structure of the variables. The method may, however, be useful in studies with large n as a preliminary analysis to locate the most promising predictors of a criterion.

General and Key Cluster Analysis (*Bifactor and Second-order Factor Analysis*) (Table 1, column 6)

Historically, the urge to discover one general dimension in cognitive behaviors provided the impetus to the ultimate development of multidimensional analysis. It produced the two-factor theory of Spearman [29, 30], and its subsequent generalization by Holzinger and Harman [13] to bifactor analysis.

Applying domain sampling principles to this case, one defines the first dimension as a general domain score on a full battery of all n variables; thereafter each dimension is a residual score on an empirically discerned key cluster. In Table 1, column 6, notice that in bifactor analysis the decisions follow the same pattern as in CC analysis, except for the one particular of the definition of the dimensions, row d. Here, the deviation is *only* with respect to the first dimension C_1 which is defined in general formula (9) as the sum of scores on all n variable domains, i.e., $s_1 = n$. In the first residual correlation matrix and thereafter the regular CC procedure is carried through, each subsequent dimension being defined by s_x variables.

Orthodox Bifactor Analysis. As illustrated by Holzinger and Harman [13], the orthodox procedure is applied to predominately positive matrices that may not require reflection. The variables are preclustered into k' clusters, each consisting of variables which (in our terms) show high P^2 values with each other. One of these clusters is a general cluster, consisting of one variable drawn from each of the remaining $k' - 1$ clusters. The first dimension, C_1, is defined by this general cluster and is called a general factor, g, the first partial communalities of all n variables on it being called their squared

g-saturations. In the first residual matrix and thereafter, the dimensions are successively defined in turn as each of the residual $k' - 1$ clusters (group factors). The partial communalities are computed by approximation (32) but for the defining variables of each key cluster only. Zero partial communalities are by *fiat* assigned to variables in other clusters. This procedure is unwieldy computationally. Results from it will correspond closely to those more efficiently achieved by the CC method, modified as in Table 1, column 6, by defining the first dimension as a general cluster domain, the remaining dimensions as key clusters; preclustering is unnecessary.

Orthodox Second-order Factor Analysis. In Thurstone centroid analysis, when a generally positive matrix of correlations between primary factors results from rotation, the correlations between these first-order factors may then be subjected to a new centroid analysis on one dimension only. This dimension is termed a second-order factor. On domain sampling principles the k correlated primaries represent hypothetical oblique clusters which, unlike the oblique clusters discovered in key cluster analysis, are normally poorly defined by actual variables. The second-order factor is simply a composite general cluster domain—a battery score on the k oblique clusters. This general composite is difficult to interpret because of its vague omnibus character and because of the complex redundancy of some variables that are common to two or more primaries.

A cleaner general composite, if desired, would be secured by a CC analysis designed as described above, namely, by defining the first dimension C_1 as a composite domain of all n variables without redundancy, later dimensions as key clusters. To illustrate that such a first dimension corresponds closely to the Thurstone second-order dimension (sometimes called g), the writer compared the correlations between the 11 WAIS variables and the Thurstone second-order factor ([7], Table 5, 18–19 yr. olds) with their correlations with the C_1 dimension defined as an nonredundant general cluster domain. The P^2-value of the paired columns of correlations was unity in the second decimal place.

Second-order analysis can be extended, of course, to more than a single dimension and to matrices with positive and negative correlations between the first-order factors. This problem is more fruitfully approached, however, under higher-order composites of oblique clusters discussed in the last section of this paper (see "Designed reanalyses").

General Cluster Analysis (Centroid–Simple Structure, Principal Axes, Maximum Likelihood Factor Analysis) (Table 1, columns 7, 8, 9)

In the remaining group of methods each dimension C_x is defined successively as general battery residual scores on all n variable domains, unweighted or weighted.

Unweighted General Clusters (Centroid–Simple Structure Factor Analysis)

(Table 1, column 7). If the analyst wishes to define each dimension C_z as the total unweighted residual domain score on the n variables, he sets $s_z = n$ in (9) and all weights equal. The procedures of CC analysis are now called for down to structure analysis, as shown in Table 1, column 7. As pointed out earlier, the indiscriminate use of all n variables requires graphical or analytic rotational methods to describe the simple cluster structure. Following Thurstone, the simple-structure factorist who uses graphical methods usually does not place the oblique rotated dimensions through oblique clusters of variables but rather the rotated dimensions bound the variables. As a result, the calculation of individual's scores on these oblique dimensions or factors requires the use of the complex regression equation given in (27).

This complexity usually leads the simple-structure factorist, after having laboriously located the "underlying" primaries through rotational devices, to abandon efforts to estimate the scores of individuals on them. Thurstone recommends ([34], p. 515) that the complex regression score be replaced by computing a simple cluster score on the nearest oblique cluster—precisely the type of cluster domain that is directly located and measured by key cluster analysis.

Orthodox unweighted general cluster analysis has been fully developed by Thurstone [32, 34] and his followers under the name "multiple factor analysis." In Table 1, column 7, if one compares the orthodox procedure (2nd cell entries) originally formulated on the basic factor theorem with the procedures based on domain sampling (1st cell entries), they are seen to be identical except for certain unrefined features of the orthodox method: the use of highest r's as initial estimates of communalities, the terminating of factoring by a statistical test of residual r's, and the lack of reiteration of factoring to converged communalities.

Weighted General Clusters (Principal Axes or Components) (Table 1, column 8). One may wish to attach differential weights to scores on the n variable domains in (9) in order to maximize the sum of partial communalities on each successive dimension. The definition of the dimensions in this case is otherwise identical to that in the preceding unweighted case. Procedurally (see Table 1, column 8), no initial reflection of variables is required, but an electronic computer is essential to solve reiteratively by least squares for the values of the partial communalities, row e.

In the orthodox method of weighted general clusters as described by Pearson [25], Hotelling [14, 15], and Kelley [20], considerable confusion has been introduced by their use of unities (and sometimes reliabilities) in the diagonals. A rationale for unity diagonals would be equally applicable to any of the other methods described above and equally inappropriate, for communalities are called for, as stated earlier in this paper (see also [13], ch. 7). The only basic feature that distinguishes the weighting of general clusters from the preceding centroid analysis is the decision to apply weights in the definition of each dimension. A main feature that distinguishes both approaches from the key cluster method is the decision to set s_x equal to n. As a consequence both general cluster methods require rotational procedures in order to describe cluster structure.

Maximum Likelihood General Clusters (Table 1, column 9). As with the preceding two methods, Lawley's maximum likelihood procedures [21, 22, 23, 27] determine k general cluster dimensions, $s_x = n$ in (9). But Lawley's dimensions are significant in the sense that the final partial and total communalities reiteratively determined from trial values are model population values that would produce a correlation matrix from which there is maximum likelihood that the observed statistical matrix represents sampling fluctuations. A chi square test of residual correlations is taken as evidence that the model cannot be rejected. The procedures are onerous even for an electronic calculator, since models with k progressively increased must be tested one at a time. Efficiency is greatly improved if good trial values of k and of the partial communalities can be found to initiate the reiterative procedures. This maximum likelihood approach has special appeal because of its capabilities of significance testing, but its gargantuan computation requirements and the need of rotation to a final cluster structure are serious limitations.

To sum up the three general cluster approaches and offer a prospectus, all three general methods define dimensions as indiscriminate composite domain scores on all n variables. This definition is a common limitation because it leads to the uncertain procedures of rotation to simple structure. Paradoxically, the Thurstonian unweighted case, the most generally used method of factor analysis [6, 8, 10, 34], is least justifiable. A few books, notably British ones, do put the method in the right perspective [1, 3, 31]. Centroid factoring is not as exact as the principal axes method in determining dimensionality. It lacks a test of significance, unlike the maximum likelihood method. For a general cluster solution the analyst would now use a modern computer programmed for the principal axes or maximum likelihood methods. As for the computational simplicity of the centroid method, the key cluster methods are simpler and have the additional merit of routinely describing oblique structure without the need of rotational devices.

Designing principal axes and maximum likelihood solutions so as to describe oblique cluster structure would remove their present inadequacies. Coupling key cluster analysis to them would, it would seem, turn the trick. A preliminary pivot variable (PV) analysis would quickly demarcate the desired set of k oblique clusters. If a method of least squares fit of k axes through these oblique subcentroids can be devised, the result would be a weighted key cluster or principal cluster solution. To increase the efficiency of a maximum likelihood method a preliminary PV analysis should, it would seem, provide good initial trial values, not only of dimensionality but also of partial communalities (appropriately unaugmented). If this method can be further modified to pass hypothetical population dimensions through the k key clusters, rotation would not be required, but it would retain the significance testing features of Lawley. The result would be a maximum likelihood key cluster solution.

Rationally Designed Dimensional Analysis

The order in which clusters are selected to define successive dimensions is determined in the key cluster methods of Table 1, columns 1 to 4, completely objectively; the object is to maximize the independence of the variables that define different dimensions and to select them in decreasing order of salience. But key cluster analysis need not be so blindly empirical; it may be designed to test hypotheses based on theories about the structure and order of variance determination among the n variables.

Designed TC and CC Analysis. An analyst may generate the hypothesis, for example, that clusters among n_1 variables of type A (say, a group of sociological variables) may better predict the communality variances of all clusters among n_2 variables of type B (say, a group of social attitude variables), than would be the case in the converse direction, B to A. To test the hypothesis, he would perform a TC or CC analysis of the $n_1 + n_2$ matrix of correlations, but restrict the defining variables of the factored dimensions solely to the A block of variables. After the terminating criterion T has been met by the A-variables, the residual communalities of the B-variables would be their communality variances unpredictable by the A set. A fresh TC or CC analysis in reverse order, B to A, would reveal residual communality variances of the A-variables unpredictable by the B set. Under the hypothesis, percentage determination of the variances in the A to B design should be higher than in B to A.

Designed PCC and RCC Analysis. Having empirically or rationally preclustered n variables, for example, of the sociological and attitude types, an analyst may on the basis of theory generate the hypothesis that only certain ones of the k' clusters are salient predictors of the remaining clusters, and in a hypothetical order of salience. His test would consist of a dimensional analysis in which the successively factored dimensions would be defined by the different selected clusters arranged in the order of their hypothetical decreasing salience. After the T-criterion is met, then under the hypothesis the analyst would discover that, for the variables not in the selected clusters, their residual communalities should in general have progressively decreased as factoring proceeded and should have become negligible on dimensions following those hypothesized as being salient.

Designed Reanalyses. Any of the above designs may follow upon a preliminary purely empirical key cluster analysis. Such a preliminary study, including the oblique structure analysis, may lead to new ideas that the analyst may wish to test by additional types of dimensional analyses.

For example, will a small number of *higher-order*, or composite, clusters maximally predict the communality variances of the n variables? To illustrate, a study of the oblique structure of sociological and social attitude clusters may suggest that one composite of sociological clusters and one composite of attitude clusters may leave but a minor amount of the communalities of the variables unpredictable. The test in this case would be a new dimensional analysis in which the first two dimensions would be defined respectively by the two composite clusters. The residual communalities of the variables would constitute the parts unpredictable from these two dimensions. Thurstone's single second-order factor analysis is a special design, discussed earlier, in which *one* such general composite cluster may be so tested.

Another type of reanalysis consists of bringing together in one common new dimensional analysis clusters found to be the most nearly independent sets in different prior analyses. To illustrate, using urban neighborhoods as objects, the writer performed three separate CC analyses: on sociological variables in 1940, on the same characteristics in 1950, and on voting variables in 1954. Each of the separate analyses yielded three salient oblique dimensions. The nine oblique clusters were then projected into a single common dimensional analysis. This master analysis still yielded three salient oblique dimensions, demonstrating thereby the common tridimensionality of these characteristics over more than a decade. In such reanalyses the analyst may also wish to include brand new variables that he considers theoretically to belong to the common structure.

Summary

The general method of multidimensional analysis, designed on domain sampling principles, covers as special cases all the main types of cluster and factor analysis. The different methods vary primarily in special decisions about the nature of an independent dimension. Such a dimension is defined in general as a composite score on a cluster of variable-domains. In cluster analysis terms the special methods of key cluster analysis, denoted as TC, CC, PCC, and RCC, define each dimension as a selected set or subgroup from all the n variables. CC analysis is the most generally applicable. Square root or diagonal factor analysis, called PV analysis, pivots each dimension on one central variable. Bifactor and second-order factor analysis defines the first dimension as a general cluster of all n variables, the later dimensions as key clusters. Among the general cluster methods that define each dimension as a composite of all n variables, centroid factor analysis defines it as an unweighted total cluster domain, principal axes as a weighted one, and maximum likelihood factor analysis also a weighted one but having the additional feature of supplying a technique for estimating the number of statistically significant dimensions required. The key cluster methods determine simple cluster structure as a routine aspect of the factoring process, whereas the general cluster methods require laborious rotations to determine structure. The key cluster methods can be applied blind, but they can also be designed to test hypotheses.

REFERENCES

[1] Adcock, C. J. *Factorial analysis for non-mathematicians*. Carlton: Melbourne Univ. Press, 1954.

[2] Burt, C. *The distribution and relations of educational abilities*. London: King, 1917.

[3] Burt, C. *The factors of the mind*. New York: Macmillan, 1941.

[4] Burt, C. Alternative methods of factor analysis and their relations to Pearson's method of principal axes. *Brit. J. Psychol., Statist. Sec.*, 1949, **2**, 98–121.

[5] Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

[6] Cattell, R. B. *Factor analysis*. New York: Harper, 1952.

[7] Cohen, J. The factorial structure of WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

[8] Fruchter, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

[9] Garnett, J. C. M. On certain independent factors of mental measurements. *Proc. Roy. Soc.*, 1919, A, **96**, 91–111.

[10] Guilford, J. T. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

[11] Guttman, L. A new approach to factor analysis: the radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Columbia Univ. Press, 1956.

[12] Holzinger, K. J. A simple method of factor analysis. *Psychometrika*, 1944, **9**, 257–262.

[13] Holzinger, K. J. and Harman, H. *Factor analysis*. Chicago: Univ. Chicago Press, 1941.

[14] Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.

[15] Hotelling, H. Simplified calculation of principal components. *Psychometrika*, 1935, **1**, 27–35.

[16] Kaiser, H. F. Solution for the communalities: a preliminary report. Rep. No. 5, AF 41(657)-76, Univ. Calif., Berkeley, Sept., 1956.

[17] Kaiser, H. F. Further numerical investigation of the Tryon-Kaiser solution for the communalities. Rep. No. 14, AF 41(657)-76, Univ. Calif., Berkeley, May, 1957.

[18] Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

[19] Kelley, T. L. *Crossroads in the mind of man*. Stanford: Stanford Univ. Press, 1928.

[20] Kelley, T. L. *Essential traits of mental life*. Cambridge: Harvard Univ. Press, 1935.

[21] Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc., Edinburgh*, 1940, **60**, 64–82.

[22] Lawley, D. N. The maximum likelihood method of estimating factor loadings. Ch. 21 in G. Thomson, *The factorial analysis of human ability*. (5th ed.) London: Univ. London Press, 1951.

[23] Lord, F. M. A study of speed factors in tests and academic grades. *Psychometrika*, 1956, **21**, 31–50.

[24] Neuhaus, J. O. and Wrigley, C. The quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

[25] Pearson, K. On lines and planes of closest fit to systems of points in space, *Phil. Mag.*, 1901, 6th Ser., 559ff.

[26] Pinzka, C. and Saunders, D. R. Analytic rotation to simple structure. II. Extension to an oblique solution. Princeton, N. J.: Educ. Test. Serv. Res. Bull., Aug., 1954.
[27] Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, **20**, 93–111.
[28] Saunders, D. R. An analytic method for rotation to orthogonal simple structure. Princeton, N. J.: Educ. Test. Serv. Res. Bull., Aug., 1953.
[29] Spearman, C. General intelligence objectively determined and measured. *Amer. J. Psychol.*, 1904, **15**, 201–293.
[30] Spearman, C. *The abilities of man.* London: Macmillan, 1927.
[31] Thomson, G. *The factorial analysis of human ability.* (5th ed.) London: Univ. London Press, 1951.
[32] Thurstone, L. L. *The vectors of mind.* Chicago: Univ. Chicago Press, 1935.
[33] Thurstone, L. L. A multiple group of factoring the correlation matrix. *Psychometrika*, 1945, **10**, 73–78.
[34] Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. Chicago Press, 1947.
[35] Thurstone, L. L. Note about the multiple group method. *Psychometrika*, 1949, **14**, 43–45.
[36] Tryon, R. C. A theory of psychological components—an alternative to "mathematical factors." *Psychol. Rev.*, 1935, **42**, 425–454.
[37] Tryon, R. C. *Cluster analysis.* Ann Arbor, Mich.: Edwards, 1939.
[38] Tryon, R. C. Identification of social areas from cluster analysis. *Univ. Calif. Publ. Psychol.*, 1955, **8**, No. 1, 1–100. Berkeley: Univ. Calif. Press.
[39] Tryon, R. C. Reliability and behavior domain-validity: reformulation and historical critique. *Psychol. Bull.*, 1957, **54**, 229–249.
[40] Tryon, R. C. Communality of a variable: formulation from cluster analysis. *Psychometrika*, 1957, **22**, 241–259.
[41] Tryon, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt*, 1958, **18**, 3–35.
[42] Tryon, R. C. General dimensions of individual differences: cluster analysis vs. multiple factor analysis. *Educ. psychol. Measmt*, 1958, **18**, 477–495.
[43] Wrigley, C. F. and Neuhaus, J. O. The matching of two sets of factors. *Amer. Psychologist*, 1955, **10**, 418–419. (Abstract)
[44] Wrigley, C. F., Cherry, C. N., Lee, M. C., and McQuitty, L. L. Use of the square root method to identify factors in the job performance of aircraft mechanics. *Psychol. Monogr.*, 1956, **71**, No. 1 (Whole No. 430).
[45] Wrigley, C. The effect upon the communalities of changing the estimate of the number of factors. Rep. No. 13, AF 41(657)-76, Univ. Calif., Berkeley, March, 1957.
[46] Wrigley, C. F. An empirical comparison of various methods for estimating communalities. *Educ. psychol. Measmt*, in press.

Manuscript received 5/7/58

Revised manuscript received 9/10/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

LEAST SQUARES ESTIMATION IN FINITE MARKOV PROCESSES

ALBERT MADANSKY

RAND CORPORATION

The usual least squares estimate of the transitional probability matrix of a finite Markov process is given for the case in which, for each point in time, only the proportions of the sample in each state are known. The purpose of this paper is to give another estimate of this matrix and to investigate the properties of this estimate. It is shown that this estimate is consistent and asymptotically more efficient than the previously considered estimate in a sense defined in this paper.

Notation

The matrix conventions and terminology used in [2, 4, and 6] will be followed. Let m_{ik} $(i = 1, 2, \cdots, a; k = 1, 2, \cdots, n)$ be the proportion observed on trial k in alternative category i of a multinomial population based on a sample of size S. Let $E(m_{ik}) = \mu_{ik}$, where $0 \leq \mu_{ik} \leq 1$ and $\sum_i \mu_{ik} = 1$ for each k. Let

$$M = \begin{bmatrix} m_{11} & \cdots & m_{1,n-1} \\ \vdots & & \vdots \\ m_{a1} & \cdots & m_{a,n-1} \end{bmatrix}.$$

Also, let t_{ij} be the transitional probability that an observation which is in alternative category j at a given trial be in alternative category i at the next trial $(i, j = 1, 2, \cdots, a)$. Define $T_i = [t_{i1} \cdots t_{ia}]$ and

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1a} \\ \vdots & & \vdots \\ t_{a1} & \cdots & t_{aa} \end{bmatrix}.$$

Let $N_i = [m_{i2} \cdots m_{in}]$, and let N be the $a \times (n - 1)$ matrix with N_i as row i. Also let

$$\mu = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1,n-1} \\ \vdots & & \vdots \\ \mu_{a1} & \cdots & \mu_{a,n-1} \end{bmatrix},$$

and $\nu_i = [\mu_{i2} \cdots \mu_{in}]$, and let ν be the $a \times (n - 1)$ matrix with ν_i as row i.

Then $E(M) = \mu$, $E(N_i) = \nu_i$, and $E(N) = \nu$. Also, by the Markovian assumption, $\nu_i = T_i\mu$, so that $\nu = T\mu$ and $T = \nu\mu'(\mu\mu')^{-1}$ if $\mu\mu'$ is nonsingular.

One should note that since the Markov process has $a(a - 1)$ parameters all these parameters cannot be estimated unless $n(a - 1) \geq a(a - 1)$, i.e., unless the number of independent m_{ik} is greater than or equal to the number of parameters to be estimated. It therefore should be assumed that $n \geq a$. It shall, however, be assumed that $\mu\mu'$ is nonsingular and hence, the necessary condition that $\mu\mu'$ be nonsingular, that $n - 1 \geq a$. Other assumptions will be introduced as they become necessary.

Estimation of T

Since M is not a matrix of constants, the usual proof in the theory of least squares (cf. [5], p. 55) does not apply in this case to show that

$$\bar{T}_i = N_i M'(MM')^{-1}$$

is the least squares estimate of T_i . One can work conditionally on M and find the $\bar{T}_i$ which minimizes

$$(T_i M - N_i)(T_i M - N_i)' = C_i C_i' ,$$

given M. In that case

$$\bar{T}_i = N_i M'(MM')^{-1}$$

if MM' is nonsingular, as is shown in [2], and the estimate of T, *conditional on* M, which minimizes $C_i C_i'$ for each i is

$$\bar{T} = NM'(MM')^{-1}.$$

Even if M were a matrix of constants (or even if one works conditionally on M), $\bar{T}_i$ would still not be the "least squares" estimate of T_i . Although the elements of C_i have mean zero (this can be seen as a corollary of (2.14) of [1]), they are not uncorrelated and do not all have the same variance, σ^2.* In a situation such as this, the least squares estimate of T_i is obtained by transforming C_i to

$$\tilde{C}_i = C_i \Lambda_i^{-1/2},$$

whose elements are uncorrelated and have the same variance, $\sigma^2 = 1$, and by minimizing

$$\tilde{C}_i \tilde{C}_i' = C_i \Lambda_i^{-1} C_i' ,$$

---

*I am indebted to a referee for pointing out that equal variances are not required to use the method of least squares for estimation, but only necessary for the optimal properties of least squares estimates, in particular having minimum variance (as defined, e.g., in [3]) among linear unbiased estimates, to be realized. The referee also points out that cov $(m_{jk}, C_{ik}) = 0$, i.e., the error is uncorrelated with the independent variables in the relation $m_{i,k+1} = \sum_j t_{ij}\, m_{jk} - C_{ik}$. This can be seen as a direct application of Lemma 1.

where Λ_i is the covariance matrix of C_i. Let us first determine Λ_i and Λ_i^{-1}. Use is made of the following lemma.

LEMMA 1. Cov $(m_{in}, m_{r,n-k}) = \mu_{r,n-k}\,(t_{ir}^{(k)} - \mu_{in})/S$, where $t_{ir}^{(k)}$ is element (i, r) of T^k.

PROOF. Define

$$Y_{ijs} = \begin{cases} 1 \text{ if individual } s \text{ is in category } i \text{ at time } j, \\ 0 \text{ otherwise.} \end{cases}$$

Then

$$\begin{aligned} \operatorname{cov}(m_{in}, m_{r,n-k}) &= \sum_{s=1}^{S} \operatorname{cov}(Y_{ins}, Y_{r,n-k,s})/S^2 \\ &= [\Pr\{Y_{ins} = 1,\ Y_{r,n-k,s} = 1\} - \mu_{in}\mu_{r,n-k}]/S \\ &= [\Pr\{Y_{ins} = 1 \mid Y_{r,n-k,s} = 1\}\mu_{r,n-k} - \mu_{in}\mu_{r,n-k}]/S \\ &= (t_{ir}^{(k)}\mu_{r,n-k} - \mu_{in}\mu_{r,n-k})/S. \qquad \text{QED.} \end{aligned}$$

When $k = 0$, $t_{ir}^{(k)}$ is defined to be $\delta_{ir} = \begin{cases} 1 & \text{if } i = r, \\ 0 & \text{otherwise} \end{cases}$; the usual multinomial variances and covariances result, namely

$$\operatorname{cov}(m_{in}, m_{rn}) = -\mu_{in}\mu_{rn}/S$$

and

$$\operatorname{var} m_{in} = \mu_{in}(1 - \mu_{in})/S.$$

Let C_{ik} be the kth element of C_i, $k = 1, \cdots, n - 1$. Then

$$\begin{aligned} S \operatorname{var} C_{ik} &= S \operatorname{var}\left(\sum_{j=1}^{a} t_{ij}m_{jk} - m_{i,k+1}\right) \\ &= \mu_{i,k+1}(1 - \mu_{i,k+1}) + \sum_{j=1}^{a} t_{ij}^2\mu_{jk}(1 - \mu_{jk}) \\ &\qquad - \sum_{j \neq j'}^{a} t_{ij}t_{ij'}\mu_{jk}\mu_{j'k} - 2\sum_{j=1}^{a} t_{ij}\mu_{jk}(t_{ij} - \mu_{i,k+1}) \\ &= \mu_{i,k+1}(1 + \mu_{i,k+1}) - \sum_{j=1}^{a} t_{ij}^2\mu_{jk} - \left(\sum_{j=1}^{a} t_{ij}\mu_{jk}\right)^2 \\ &= \mu_{i,k+1} - \sum_{j=1}^{a} t_{ij}^2\mu_{jk}, \end{aligned}$$

and when $k > l$

$$S \operatorname{cov}(C_{ik}, C_{il}) = \mu_{i.l+1}(t_{ii}^{(k-l)} - \mu_{i.k+1}) - \sum_{j=1}^{a} t_{ij}\mu_{jl}(t_{ij}^{(k+1-l)} - \mu_{i.k+1})$$

$$- \sum_{j=1}^{a} t_{ij}\mu_{i.l+1}(t_{ji}^{(k-l-1)} - \mu_{jk})$$

$$+ \sum_{j=1}^{a} t_{ij} \sum_{r=1}^{a} t_{ir}\mu_{rl}(t_{jr}^{(k-l)} - \mu_{jk})$$

$$= -\sum_{j=1}^{a} t_{ij}[\mu_{jl}t_{ij}^{(k+1-l)} + \mu_{i.l+1}t_{ji}^{(k-l-1)}]$$

$$+ \sum_{r=1}^{a} t_{ir}\mu_{rl} \sum_{j=1}^{a} t_{ij}t_{jr}^{(k-l)} + \mu_{i.l+1}t_{ii}^{(k-l)} = 0,$$

since

$$\sum_{j=1}^{a} t_{ij}t_{jr}^{(k-l)} = t_{ir}^{(k+1-l)}.$$

Hence Λ_i is a diagonal matrix with diagonal elements

$$\mu_{i.k+1} - \sum_{j=1}^{a} t_{ij}^2\mu_{jk} .$$

Now Λ_i is an unknown matrix, but each μ is consistently estimated by its corresponding m. Also, since

$$\operatorname*{plim}_{S\to\infty} N = \nu \quad \text{and} \quad \operatorname*{plim}_{S\to\infty} M = \mu,$$

it follows that

$$\operatorname*{plim}_{S\to\infty} \bar{T} = \nu\mu'(\mu\mu')^{-1} = T,$$

so that elements of $\bar{T}$ are consistent estimates of corresponding elements of T. Therefore, $\hat{\Lambda}_i$, a consistent estimate of Λ_i , can be formed by substituting the m's and the elements of $\bar{T}$ for the corresponding μ's and t's in the above equations for the elements of Λ_i .

Consider the estimate

$$\tilde{T}_i = N_i\hat{\Lambda}_i^{-1}M'(M\hat{\Lambda}_i^{-1}M)^{-1}$$

(if $M\hat{\Lambda}_i^{-1}M'$ is nonsingular) which would be the least squares estimate of T_i if $\hat{\Lambda}_i^{-1} = \Lambda_i^{-1}$ and if Λ_i were not a function of the t's. It is easy to see that $\tilde{T}_i$ is also a consistent estimate of T_i .

If R and Q are symmetric matrices then $R \geq Q$ if and only if $x(R - Q)x' \geq 0$ for all vectors x. Let us adapt the definition of minimum variance linear unbiased estimate of a vector-valued parameter given in ([3], section 2.5) as follows. It may be said that of two consistent estimators, say u and v, of a vector-valued parameter θ, u is asympotically minimum variance if the asymptotic covariance matrix of $\sqrt{S}\,(u - \theta)$ is less than or equal to the asymptotic covariance matrix of $\sqrt{S}\,(v - \theta)$ in the sense defined above.

THEOREM. *Let* $\tilde{\Sigma}_i$ *be the asymptotic covariance matrix of* $\sqrt{S}\,(\tilde{T}_i - T_i)$ *and let* $\bar{\Sigma}_i$ *be the asymptotic covariance matrix* $\sqrt{S}\,(\bar{T}_i - T_i)$. *Then* $\tilde{\Sigma}_i \leq \bar{\Sigma}_i$.

PROOF. We know that

$$m_{kl} = \mu_{kl} + O(1/\sqrt{S}),$$

since for fixed l, m_{kl} is the maximum likelihood estimate of μ_{kl}. Therefore

$$\bar{t}_{ij} = t_{ij} + O(1/\sqrt{S})$$

and also

$$\tilde{t}_{ij} = t_{ij} + O(1/\sqrt{S})$$

for all i and j. Then write

$$N_i = \nu_i + X/\sqrt{S}, \qquad M = \mu + Y/\sqrt{S}, \quad \text{and} \quad \hat{\Lambda}_i^{-1} = \Lambda_i^{-1} + Z/\sqrt{S}.$$

To terms of order $1/\sqrt{S}$,

$$MM' = \mu\mu' + (\mu Y' + Y\mu')/\sqrt{S},$$

so that

$$(MM')^{-1} = [I - (\mu\mu')^{-1}(\mu Y' + Y\mu')/\sqrt{S}](\mu\mu')^{-1}$$

to terms of order $1/\sqrt{S}$. Similarly,

$$(M\hat{\Lambda}_i^{-1}M')^{-1} = [I - (\mu\Lambda_i^{-1}\mu')^{-1}(\mu\Lambda_i^{-1}Y' + \mu Z\mu' + Y\Lambda_i^{-1}\mu')](\mu\Lambda_i^{-1}\mu')^{-1}$$

to terms of order $1/\sqrt{S}$. Therefore, asymptotically,

$$\sqrt{S}\,(\bar{T}_i - T_i) = [(\nu_i Y' + X\mu') - T_i(\mu Y' + Y\mu')](\mu\mu')^{-1}$$

and

$$\sqrt{S}\,(\tilde{T}_i - T_i) = [(\nu_i Z\mu' + X\Lambda_i^{-1}\mu' + \nu_i\Lambda_i^{-1}Y') \\ - T_i(\mu\Lambda_i^{-1}Y' + \mu Z\mu' + Y\Lambda_i^{-1}\mu')](\mu\Lambda_i^{-1}\mu')^{-1}.$$

Since $\nu_i = T_i\mu$, the above equations reduce to

$$\sqrt{S}\,(\bar{T}_i - T_i) = (X - T_iY)\mu'(\mu\mu')^{-1}$$

and

$$\sqrt{S}\,(\tilde{T}_i - T_i) = (X - T_iY)\Lambda_i^{-1}\mu'(\mu\Lambda_i^{-1}\mu')^{-1}.$$

But the covariance matrix of $X - T_iY$ is Λ_i, so that

$$\bar{\Sigma}_i = (\mu\mu')^{-1}(\mu\Lambda_i\mu')(\mu\mu')^{-1}$$

and

$$\tilde{\Sigma}_i = (\mu\Lambda_i^{-1}\mu')^{-1}.$$

By using the Schwarz inequality for matrices given in ([3], section 2.5), it follows immediately that

$$\bar{\Sigma}_i \geq \tilde{\Sigma}_i \ .$$

Thus $\tilde{T}_i$ is asymptotically more efficient, in the sense defined above, than is $\bar{T}_i$. In fact, a modification of the above proof will show that $\tilde{T}_i$ is asymptotically at least as efficient as any other estimator of the form $N_iPM'(MPM')^{-1}$, where P is a positive definite matrix.

One should note that $\tilde{T}$ is obtained by first computing $\bar{T}$ and then modifying it. One can just as well, via the same procedure as above, modify $\tilde{T}$ to obtain another estimate of T, T^*, say. However, the asymptotic efficiency of T^*_i will be the same as that of $\tilde{T}_i$ for each i, since T^*_i will be of the form $N_iPM'(MPM')^{-1}$, where P is $\hat{\Lambda}_i^{-1}$ based on $\tilde{T}$, i.e.,

$$P = \Lambda_i^{-1} + O(1/\sqrt{S}).$$

It has been shown in [2] that $\bar{T}$ has the desirable property that $e\bar{T} = e$, where e is the a-dimensional vector $(1\ 1\ \cdots\ 1)$. That this is not true of $\tilde{T}$ can be seen from the example of the next section.

Discussion

Let us consider Miller's example [cf. 6]. In his case $a = 2$, $n = 20$, and

$$\bar{T} = \begin{bmatrix} .92 & .39 \\ .08 & .61 \end{bmatrix}.$$

Using this estimate and N and M, the diagonal elements of $\hat{\Lambda}_1$ are found to be

$$[.20\ -.04\ -.03\ .23\ .09\ .09\ -.01\ .16\ .19\ .02$$
$$-.01\ .16\ .19\ .22\ .15\ -.05\ .09\ .19\ .12]$$

and the diagonal elements of $\hat{\Lambda}_2$ are

$$[.11\ .28\ .25\ .05\ .12\ .12\ .22\ .08\ .02\ .16$$
$$.22\ .08\ .02\ -.04\ -.01\ .19\ .12\ .02\ .06].$$

Note that some of these elements are negative. However, since these elements are estimates of variance (which are non-negative quantities) these "estimates" were replaced by .001 (zero cannot be used since $\hat{\Lambda}_i$ must be nonsingular). It is then found that

$$\tilde{T} = \begin{bmatrix} .81 & .25 \\ -.02 & .49 \end{bmatrix}.$$

This example illustrates that $e\tilde{T} \neq e$ in general. It also shows that elements of $\tilde{T}$ may be inadmissible in that they can turn out to be negative. However, as Goodman points out [2], $\bar{T}$ also has this latter fault.

We suggest in this case the estimate

$$\bar{\bar{T}} = \begin{bmatrix} .81 & .25 \\ .19 & .75 \end{bmatrix}.$$

To compare $\bar{\bar{T}}$ with $\bar{T}$, it suffices to compare the covariance matrices of $\bar{\bar{T}}_1 = \tilde{T}_1$ and $\bar{T}_1$ (since $\bar{\bar{T}}_2 = e - \tilde{T}_1$ and $\bar{T}_2 = e - \bar{T}_1$). An estimate of $\tilde{\Sigma}_1$ is $(M\hat{\Lambda}_1^{-1}M')^{-1}$, and $\bar{\Sigma}_1$ is estimated by

$$(MM')^{-1}(M\hat{\Lambda}_1M')(MM')^{-1}.$$

After re-estimating Λ_1 by using $\bar{\bar{T}}$ (and finding that all the diagonal elements of this estimate are positive), we used this estimate and found that

$$\text{est } \tilde{\Sigma}_1 = \begin{bmatrix} .049 & -.123 \\ -.123 & .517 \end{bmatrix}$$

and

$$\text{est } \bar{\Sigma}_1 = \begin{bmatrix} .059 & -.149 \\ -.149 & .616 \end{bmatrix}.$$

The easiest way to see that est $\bar{\Sigma}_1 \geq$ est $\tilde{\Sigma}_1$, where $\geq$ is as defined above, is to note that all principal minors of est $\bar{\Sigma}_1 -$ est $\tilde{\Sigma}_1$ are non-negative.

Goodman also states in [2] that if the observed transitional proportions are available, they would clearly be more appropriate in the estimation of transitional probabilities. In [1] Anderson and Goodman give estimates of T and the asymptotic variances and covariances of the elements of their estimate when the observed transitional proportions are available. Calling their estimates $\hat{t}_{ij}$, the asymptotic variances and covariances may be written as

$$\text{var } \hat{t}_{ij} = t_{ij}(1 - t_{ij})/\phi_i, \qquad \text{cov } (\hat{t}_{ij}, \hat{t}_{i'j'}) = -\delta_{jj'}t_{ij}t_{i'j'}/\phi_i,$$

where

$$\delta_{ik} = \begin{cases} 1 & i = k \\ 0 & i \neq k, \end{cases}$$

and

$$\phi_i = \sum_{j=1}^{a} \sum_{k=2}^{n} \mu_{j1}t_{ii}^{[k-1]} = \sum_{k=2}^{n} \mu_{ik}.$$

We can estimate μ_{ik} by m_{ik} and so, in the example under consideration, we can estimate $\hat{\Sigma}_1$, the asymptotic covariance matrix of the estimate of T_1 (and hence $T_2 = e - T_1$) when the observed transitional proportions are available, in order to see the improvement when this information is available. It is found that est $\phi_1 = 15.3$ and est $\phi_2 = 3.7$ so that

$$\text{est } \hat{\Sigma}_1 = \begin{bmatrix} .010 & .000 \\ .000 & .051 \end{bmatrix},$$

using $\bar{\bar{T}}$ to estimate T. Hence est $\hat{\Sigma}_1 \leq$ est $\tilde{\Sigma}_1$ in this case.

With regard to the amount of extra computation involved in computing $\tilde{T}$, one can see from the form of the estimate that once $\bar{T}$ is given, the modification process will take about as much time as the computation of $\bar{T}$ itself. Since we have no idea of the relative decrease in the variances of the estimates, we cannot discuss the trade-off between the doubled computation time and the reduction in variance. This trade-off is, however, an important practical factor in determining whether $\tilde{T}$ or $\bar{T}$ is used.

REFERENCES

[1] Anderson, T. W. and Goodman, L. A. Statistical inference about Markov chains. *Ann. math. Statist.*, 1957, **28**, 89–110.

[2] Goodman, L. A. A further note on "Finite Markov processes in psychology." *Psychometrika*, 1953, **18**, 245–248.

[3] Grenander, U. and Rosenblatt, M. *Statistical analysis of stationary time series.* New York: Wiley, 1957.

[4] Kao, R. C. Note on Miller's "Finite Markov processes in psychology." *Psychometrika*, 1953, **18**, 241–243.

[5] Kempthorne, O. *The design and analysis of experiments.* New York: Wiley, 1952.

[6] Miller, G. A. Finite Markov processes in psychology. *Psychometrika*, 1952, **17**, 149–167.

Manuscript received 6/13/58

Revised manuscript received 11/8/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

AN AUGMENTED MODEL FOR SPONTANEOUS REGRESSION AND RECOVERY*

DAVID McCONNELL†

INDIANA UNIVERSITY

The Estes model for spontaneous recovery and regression is modified by relaxing the assumption that within-experiment stimulus fluctuation can be treated as negligible in the general case. The augmented model allows description of stimulus fluctuation both between and within experimental periods. Application of the model to mock experimental situations illustrates some of its properties.

This paper is based upon experimental findings reported in [5], and represents an effort to broaden the scope of Estes' model for spontaneous regression and recovery [2] so that it may hold for a larger class of data. In the original model, Estes partitions the total population S^* of stimulus elements available in the experimental situation into two sets S and S'. S is the set available during an experimental period, and S' is the set unavailable during an experimental period. The numbers of elements in the sets are related by the expression $N + N' = N^*$. The probability j, defined on S, is the probability that an element selected at random will leave S and enter S' during an instant Δt of a rest period between experimental periods. Similarly, j' is the probability that an element from S' will enter S during that instant.

These probabilities are undefined by Estes during experimental periods. In the augmented model they will be defined both within and between experimental periods. As in the Estes model, the equation

$$\frac{N}{N + N'} = \frac{j'}{j + j'} = J$$

relates these probabilities to the respective subset sizes. The augmented model thus retains the basic notation of the Estes model. It will, however, be more convenient to regard S as the set in which direct conditioning or extinction of stimulus elements is possible, and S' as the set in which no direct conditioning is possible. From period to period, and during rest periods between experimental periods, S^* is considered to remain unchanged, i.e., the same elements will be contained in S^* and their number N^* will be fixed.

---

*This study is based upon a dissertation submitted to the Department of Psychology, Indiana University, in partial fulfillment of the requirements for the Ph.D. degree.

†Now at Ohio State University.

The values of j and j' may vary from intraperiod to interperiod intervals, but will be assumed constant, on the average, within any one interval. Lastly, the proportions of conditioned elements in S^*, S, or S' will be considered to remain unaltered when values of j or j' shift.

The general situation requires a stimulus-sampling operator which will carry out conditioning of new elements in S for any values of the parameters j, j'. This operator, ϕ, must be such that the proportion of conditioned elements at the conclusion of trial $n + 1$ is related to the proportion on the previous trial by an expression such as $P_{n+1} = \phi P_n$. An operator invoked some time ago by Estes and Burke [4], such that $P_{n+1} = P_n + \theta(1 - P_n)$, (where θ is the proportion of elements from S sampled by the organism on trial $n + 1$) will carry out conditioning in S when j and j' are both 0, but cannot apply unmodified when the exchange parameters are nonzero. It should be observed that the applications made by these investigators involve what Bush and Mosteller [1] call "experimenter-controlled" reinforcement situations, while extension of the operator to the present experimental context [5] entails "subject-controlled" reinforcement. This distinction, at least in the case of animal conditioning, may be considered that between respondent conditioning and operant conditioning.

The operator ϕ will be applied in the subset S exclusively. Response probabilities in S' and in S, respectively, immediately after a trial on which ϕ is applied will be related to probabilities immediately after the previous trial by the following equations.

$$P'_{n+1} = P'_n(1 - j') + P_n jN/N', \tag{1'}$$

and

$$P_{n+1} = P_n(1 - j) + P'_n j'N'/N + \theta[1 - P_n(1 - j) - P'_n j'N'/N]. \tag{2'}$$

Equation (1′) states that the proportion of conditioned elements in S' at the conclusion of trial $n + 1$ is equal to the proportion $(1 - j')$ of those found in S' at the conclusion of the preceding trial, plus the weighted proportion jN/N' of those found in S at the conclusion of the preceding trial (the latter having escaped from S into S' during the intertrial interval). Equation (2′) states that the proportion of conditioned elements in S at the conclusion of trial $n + 1$ is equal to the proportion $(1 - j)$ of those found in S at the conclusion of the preceding trial, plus the weighted proportion $j'N'/N$ of those in S' at the same time, plus the proportion θ of the unconditioned elements from the same sources (these having been sampled and conditioned on trial $n + 1$). Rewriting (1′) and (2′),

$$P'_{n+1} - (1 - j')P'_n - j'P_n = 0, \tag{1}$$

and

$$P_{n+1} - (1 - \theta)(1 - j)P_n - (1 - \theta)jP'_n = \theta. \tag{2}$$

These equations are the basic finite difference equations for the augmented model; both the exchange-free learning equation described by Estes and Burke [4] and Estes' general expression (equation (3) of [2]) for spontaneous recovery and regression between periods can be derived from them as special cases. For the first case, note that when $j = j' = 0$, (1) and (2) become

$$P'_{n+1} - P'_n = 0, \tag{1a}$$

and

$$P_{n+1} - (1 - \theta)P_n = \theta. \tag{2a}$$

These have the solutions $P'_n = P'_0$, and $P_n = 1 - (1 - P_0)\,(1 - \theta)^n$, which are identical with those of Estes and Burke [4].

For the second case, note that when j, j' are not equal to 0, but $\theta = 0$,

$$P'_{n+1} = (1 - j')P'_n + j'P_n\,, \tag{1b}$$

and

$$P_{n+1} = (1 - j)P_n + jP'_n\,. \tag{2b}$$

From (2b),

$$P'_{n+1} = \frac{P_{n+2} - (1 - j)P_{n+1}}{j}, \quad \text{and} \quad P'_n = \frac{P_{n+1} - (1 - j)P_n}{j};$$

substituting back into (1b),

$$P_{n+2} - (2 - j - j')P_{n+1} + (1 - j - j')P_n = 0.$$

This last equation has a characteristic equation whose roots are

$$r_{1,2} = \frac{2 - j - j' \mp \sqrt{(2 - j - j')^2 - 4(1 - j - j')}}{2} = 1 - j - j',\ 1.$$

Then,

$$P_n = c_1(1 - j - j')^n + c_2 = c_1a^n + c_2\,.$$

Thus

$$P_{n+1} = c_1a^{n+1} + c_2\,,$$

whence, on substituting in (2b),

$$P'_n = -\,(j'/j)\,c_1a^n + c_2\,.$$

From the equations for P_n and P'_n,

$$P_0 = c_1 + c_2\,, \quad \text{and} \quad P'_0 = (-j'/j)c_1 + c_2\,,$$

whence

$$c_1 = \frac{j}{j + j'}\,(P_0 - P'_0), \quad \text{and} \quad c_2 = P_0 - c_1\,.$$

Thus,

$$P_n = \frac{j}{j + j'} (P_0 - P_0')a^n + P_0 - \frac{j}{j + j'} (P_0 - P_0'),$$

or, since

$$\frac{j}{j + j'} = 1 - J, \quad \text{where} \quad J = \frac{j'}{j + j'},$$

this can be written

$$P_n = P_0[J + (1 - J)a^n] + P_0'(1 - J)(1 - a^n).$$

This expression is identical, when t is substituted for n, with Estes' basic equation for spontaneous recovery and regression. The corresponding equation for S' is

$$P_n' = P_0 J(1 - a^n) + P_0'[1 - J(1 - a^n)].$$

For the general case, when j, j' and θ are nonzero, note that (2) and (1) are in the forms, respectively,

(3)
$$u_{x+1} - au_x - bv_x = \theta,$$

and

(4)
$$v_{x+1} - dv_x - j'u_x = 0,$$

where $u = P$, $v = P'$, $x = n$, $a = (1 - \theta)(1 - j)$, $b = (1 - \theta)j$, $d = 1 - j'$. From (3),

$$v_{x+1} = \frac{u_{x+2} - au_{x+1} - \theta}{b}; \qquad v_x = \frac{u_{x+1} - au_x - \theta}{b}.$$

Upon substituting in (4),

(4a)
$$\frac{u_{x+2} - au_{x+1} - \theta}{b} - \frac{d}{b}(u_{x+1} - au_x - \theta) - j'u_x = 0,$$

which becomes

(4b)
$$u_{x+2} - (a + d)u_{x+1} + (ad - j'b)u_x = (1 - d)\theta,$$

or

(5)
$$P_{n+2} - (a + d)P_{n+1} + (ad - j'b)P_n = j'\theta.$$

This is in the form

(6)
$$AP_{n+2} + BP_{n+1} + CP_n = D\mu^x,$$

for which the characteristic equation is

(7)
$$\psi(r) = Ar^2 + Br + C = 0.$$

If the two roots of this equation are real and unequal, whether positive or negative, the explicit solution will have the following form.

$$P_n = \frac{D\mu^x}{\psi(\mu)} + c_1 r_1^n + c_2 r_2^n \,. \tag{8}$$

It easily can be proved that (7) can have no complex roots. It would have such roots when $B^2 < 4C$, or when, expanding, $a^2 + 2ad + d^2 < 4ad + 4j'b$, or transposing, $a^2 - 2ad + d^2 < -4j'b$, or $(a - d)^2 < -4j'b$. But $(a - d)^2$ is always positive, as are j' and $b = (1 - \theta)j$, so the expression $(a - d)^2 < -4j'b$ can never hold, and neither can the original expression $B^2 < 4C$. Thus $r_{1,2}$ must always be real.

Now,

$$\frac{D\mu^x}{\psi(\mu)} = 1, \tag{9}$$

after substitution.* Thus, from (8),

$$P_n = 1 + c_1 r_1^n + c_2 r_2^n \,. \tag{10}$$

The roots of (7) are

$$r_{1,2} = \frac{-B \mp \sqrt{B^2 - 4C}}{2}, \tag{11}$$

since $A = 1$. From this it can be shown

$$0 < r_1 < r_2 < 1, \tag{12}$$

after substitution. Therefore $P_\infty = 1$, and $P_0 = 1 + c_1 + c_2$. From (2) and (10)

$$\begin{aligned} P_n' &= \frac{1 + c_1 r_1^{n+1} + c_2 r_2^{n+1} - (1 - \theta)(1 - j)(1 + c_1 r_1^n + c_2 r_2^n) - \theta}{(1 - \theta)j} \\ &= 1 + k_1 r_1^n + k_2 r_2^n \,, \end{aligned} \tag{13}$$

where

$$k_1 = \frac{c_1(r_1 - (1 - \theta)(1 - j))}{(1 - \theta)j}, \quad \text{and} \quad k_2 = \frac{c_2(r_2 - (1 - \theta)(1 - j))}{(1 - \theta)j}. \tag{13a}$$

Then $P_\infty' = 1$, and $P_0' = 1 + k_1 + k_2$. From this,

$$c_2 = \frac{(1 - P_0)(r_1 - (1 - \theta)(1 - j)) - (1 - P_0')(1 - \theta)j}{r_2 - r_1}, \tag{10a}$$

and

$$c_1 = -(1 - P_0) - c_2 \,. \tag{10b}$$

---

*$A = 1$, $B = -(a + d)$, $C = ad - j'b$, $D = j'\theta$, $\mu = 1$, $a = (1 - \theta)(1 - j)$, $b = (1 - \theta)j$, $d = 1 - j'$.

Now, from (10) and (10a) it is evident that $P_n = QP'_0$ (where Q is a variable positive coefficient) since c_2 is linearly related to P'_0, and $P_n = Qc_2(r_2 - r_1) = Qc_2$, because of (12).

This relationship permits equal initial probabilities in the available set S to diverge as a period progresses, the higher curve obtaining for the higher starting probability in the unavailable set S'. It can also result in crossovers between initially disparate curves, such as those found in the relative frequency measures of the experiment [5] leading to this paper. Graphs of the augmented model, based on (10) and (13) above, and using dummy parameter values, appear in Figure 1.

Figure 1 depicts the application of the augmented model to a mock experiment involving the same conditions reported in the empirical paper associated with this study [5]. Although care was taken to insure that the ordering of curves in Figure 1 corresponds to that obtained in the empirical study, this was done only to demonstrate that the results of that study could fall within the domain of the model presented here. No effort was made at curve fitting—until such time as new experimental predictions are made from the augmented model, curve fitting would be at best second guessing.

The particular parameter values chosen do *not* in general uniquely determine the form of the functions shown; and these values were in some cases selected only to illustrate special points. The values of P_0 and P'_0 in period I were taken from the observed relative frequencies correct in the first minute of reinforced responding [5]. The high value of θ (.8) chosen for periods I and II is intended to illustrate that even with very rapid acquisition and extinction, reversals of initial probability relationships, such as that demonstrated in period II of the empirical study, can be expected between immediate and delayed extinction groups if intraperiod exchange between S and S' is permitted. The large value of j, together with the disproportionately small value of j' chosen, is intended to emphasize that the corresponding value of J within the two periods (.0141) may be markedly different from that prevailing between periods ($J = .8$ between periods I and II).

In period III, one sees how a radical (80-fold) reduction in θ affects terminal probability. Here, for the first time, response probability in S' actually decreases across the period for two groups—a result not superficially obvious in considering the augmented model.

The main object of the selection of parameter values for period IV is to illustrate the nonexchange learning situation for the same value of θ used in the previous period, where exchange was permitted. Note that probabilities are stationary in the unavailable set, which in the augmented model is the exception rather than the rule. In the Estes model, this condition was assumed throughout experimental periods [2].

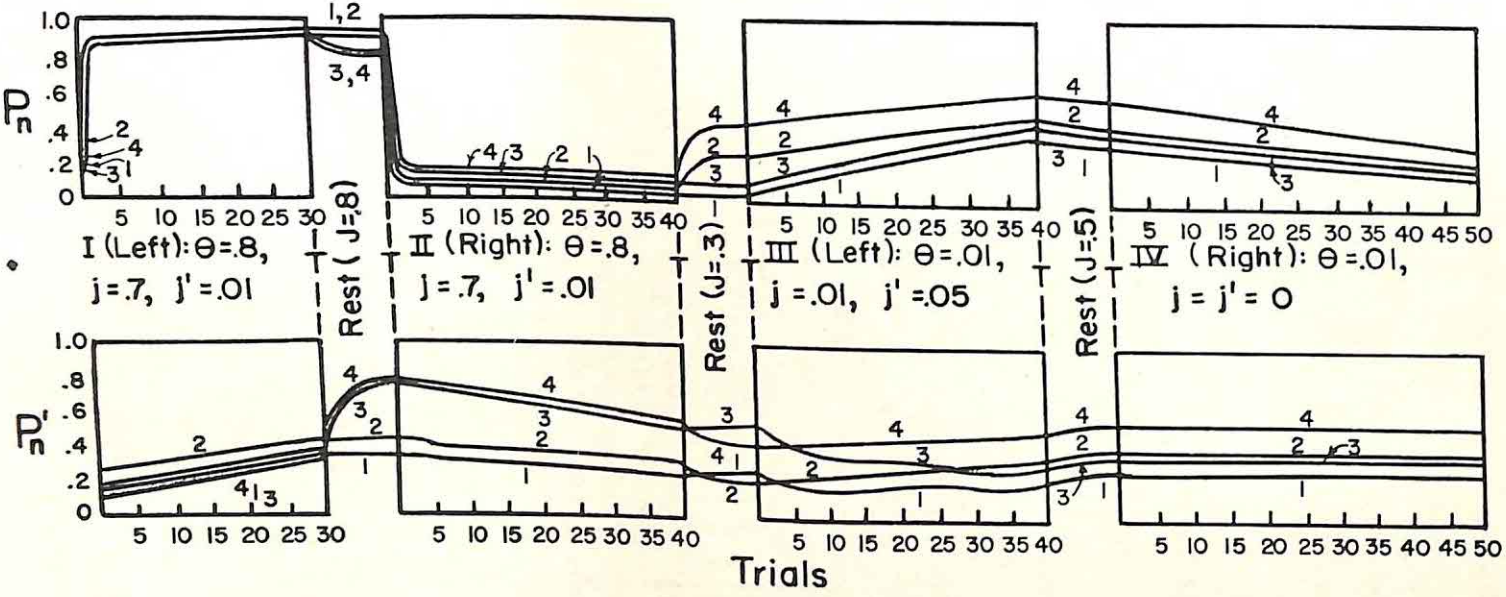


FIGURE 1

Characteristics of the augmented model using dummy values for the parameters θ, j, and j' during conditioning, and for J during the interpolated rest periods. The vertical dashed lines marked T represent transition points between different values of J. The probability plotted is that of the conditioned response.

Discrepancies Between Initial Probabilities in S and S′

Above it is shown how an excessively high proportion of conditioned elements in the set S' can retard extinction and bring about a reversal in ordering of immediate and delayed groups in a simulated experiment. Now consider the case in which $P'_0 < P_0$ at the start of an acquisition period. Referring to (2) of the augmented model, and solving for ΔP_n, gives

$$\Delta P_n = \theta(1 - P_n) + j(1 - \theta)(P'_n - P_n). \tag{14}$$

It is clear from this expression that if the factor $(P'_n - P_n)$ is negative and sufficiently large, the increment ΔP_n becomes negative. This implies that probability in the available set will decrease until such time as the condition

$$(P'_n - P_n) = \frac{\theta(1 - P_n)}{j(1 - \theta)}$$

holds, and $\Delta P_n = 0$. When θ is large and conditioning proceeds rapidly, such an initial discrepancy may produce no visible effect on the conditioning curve. But when θ is not large, there should be a detectable dip in response probability in S at the start of an acquisition period.

Figure 2 shows a low-θ dummy experiment, in which by the trial 20 the dip in available set probability has reversed. Note that after its initial intersection with the curve for the unavailable set, the available set curve increases monotonically below the other curve. These same equations may be applied to an extinction period by conceiving of the period as one of acquisition of the previously unreinforced response. We may conjecture that

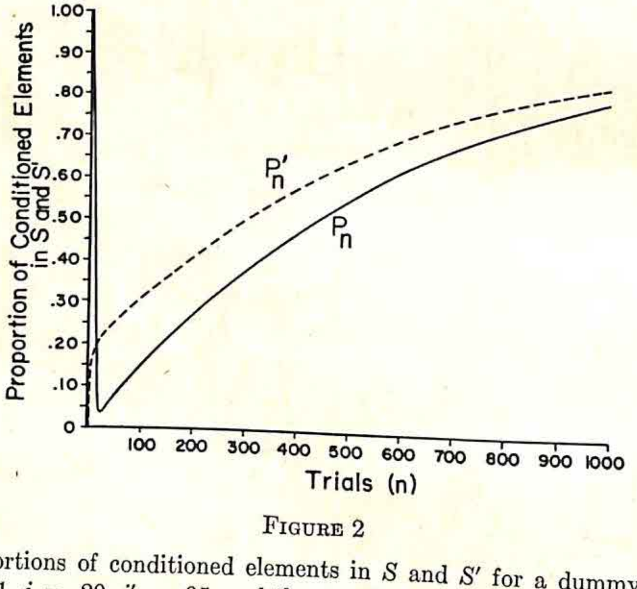


FIGURE 2

Curves of proportions of conditioned elements in S and S' for a dummy experiment in which $\theta = .01$, $j = .20$, $j' = .05$, and the starting probabilities are $P_0 = 1$, $P'_0 = 0$.

when initial probability of the response *to be extinguished* is less in the available set than is the proportion of elements conditioned to it in S', there should be detectable an initial *rise* in response probability, preceding the typically decreasing extinction function. In general, however, exaggerated discrepancies in probabilities such as could obtain at the end of an acquisition period during which values of j and j' are both low, will *not* result in these early phenomena in a subsequent period, because there is no reason to assume changes in the parameters j and j'.

Without such changes the exchange of elements will proceed no more rapidly than in the preceding period, with the result that the arrival of elements from S' will be offset by conditioning at the same rate as in the previous period. This will be true also for the extinction phenomenon. Barring the assumption of changes in the drift parameters from one period to the next, the only way to demonstrate these phenomena is to produce the necessary probability discrepancies "artificially." Consider, for example, certain elements introduced into the available set at the start of acquisition, conditioned to the reinforced response in sufficient number to raise the initial probability of the response in S markedly above the corresponding proportion of conditioned elements in S'. Then an influx of elements from S' could be expected to bring about the initial dip in question. The elements introduced might be associated with transitory phenomena such as temperature changes, handling by the experimenter, or transfer effects from other situations.

Period Spacing

Estes, still with the assumptions that $j = j' = 0$ within periods and that terminal response probability for the reinforced response in each period is 1, predicts that increasing time between periods will accelerate the accumulation of conditioned elements in S^* [3]. The augmented model permits the same prediction, but qualifies it to the extent that if j and j' are high during an experimental period, increasing the time between periods will have virtually no effect on response probability. To document this qualification, consider the extreme case of $j = j' = 1$ within experimental periods. Under this condition, the general forms of the solutions (10) and (13) still apply. However, both equations now have roots $\mp \sqrt{1 - \theta}$. The only effect of the negative root, since trials are discrete, is to introduce a slight oscillation into the growth curves of the proportions of conditioned elements in S and S'. But the outcome of the increase in j and j' is to accelerate drastically the accumulation of conditioned elements in S'. At the completion of several trials the proportions of conditioned elements would be indistinguishable from 1 in both S and S'. When, therefore, j and j' are large, so that equal limits are quickly approached by response probabilities in S and S', it is clear that increasing time between periods cannot increase the over-all rate of conditioning in S^*.

Estes also predicts that in a sequence of randomly alternated acquisition and extinction periods, response probability in S^* will approach the asymptote π, the proportion of acquisition periods in the sequence, and that this asymptote will be independent of values of J and of the time between periods. In general, predictions for mixed sequences, whether based on the augmented model or on Estes' model, must be applied cautiously. Both models result in the prediction that π will be the expected value at time t following the mth in a sequence of such periods. But while the approach of the *expectation* to π is monotonic across the sequence, it is by no means true that for a fixed experimental sequence the interperiod asymptotes must each approach π.

This point is illustrated in Fig. 3, representing four dummy experiments, for each of which the proportion of conditioned elements in both available and unavailable sets is plotted period by period. Examination of the first and third pairs of these curves demonstrates that even with the Estes assumption ($j = j' = 0$) regular oscillation can be expected to continue indefinitely for any subject or for any group whose members are identically treated—

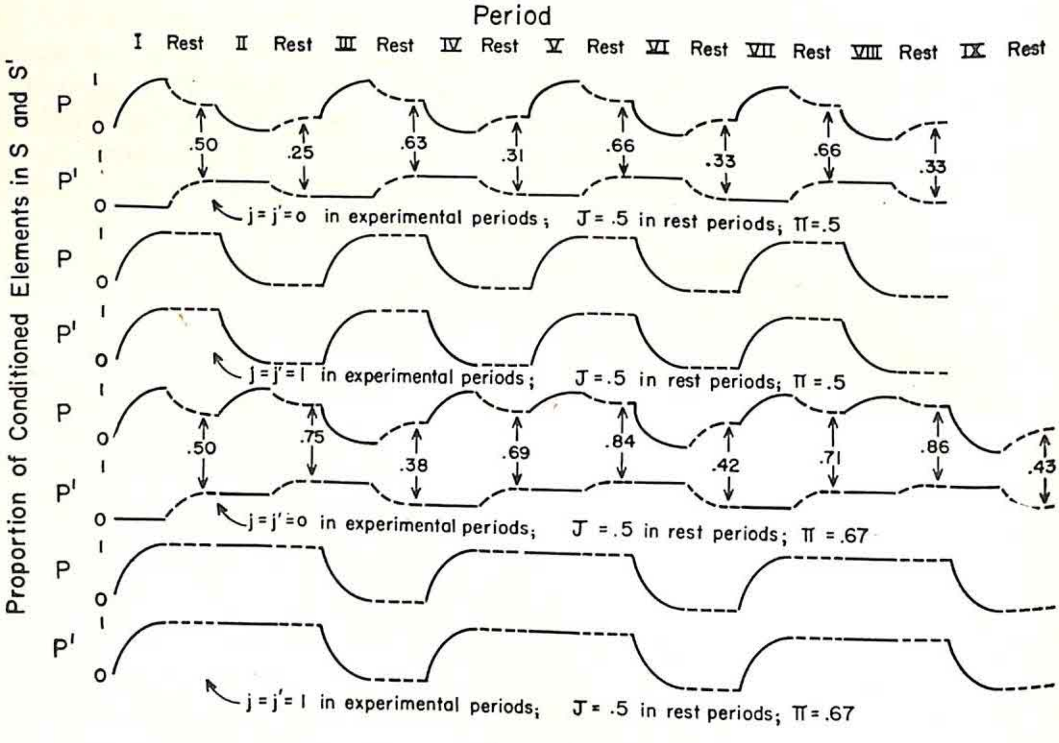


FIGURE 3

Response probability in S and S' for four dummy experiments involving a mixed sequence of acquisition and extinction periods.

in spite of the fact that the *mean* interperiod asymptote is clearly π. This oscillation should of course disappear if *Ss* are randomly assigned to different schedules and response curves are plotted as the mean of all measures at a given trial or time. The second and fourth pairs of curves in Figure 3 show what should be expected in the extreme case of the augmented model when $j = j' = 1$, and when π is the same as in the first and third experiments, respectively. It is clear that the extrema of oscillations may be significantly greater than those for the Estes model; it is evident that this should be true to some degree even for j and j' less than 1. However, it is equally clear that the mean interperiod asymptote in the extreme oscillatory cases is still π, as it is for the Estes model. The advantage of the augmented model for long, mixed sequences of periods would thus appear to accrue from its ability to interpret interperiod asymptotes which are inordinately higher or lower than the expected value of π. Since, however, some oscillation from period to period is to be expected with the Estes assumptions the intraperiod exchange notion should *not* be applied on this basis alone. The criteria of intraperiod reversals of order or nonmonotonic response functions should first be invoked, since only they are unique to the free-exchange situation.

The potential applicability of the augmented model to a variety of psychological situations appears to the author to be derived from the simple notion that stimulus fluctuation within even relatively short experimental periods can be evaluated by means of the simultaneous equations proposed, without being relegated to the status of a variable "outside" the model. It is conceded that these equations do not simplify the mathematics of Estes' basic model, but relief is in sight on this score from digital computers. A number of routines are now available for simultaneous solution of n equations using trial values of constants by use of high-speed computers such as the IBM 650 and 701 or 704. These routines are to be found in computing facility libraries at Indiana and Ohio State, and are available by request.

REFERENCES

[1] Bush, R. R. and Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

[2] Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

[3] Estes, W. K. Statistical theory of distributional phenomena in learning. *Psychol. Rev.*, 1955, **62**, 369–377.

[4] Estes, W. K. and Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

[5] Mc Connell, D. G. Spontaneous regression and recovery in a sequence of discrimination periods. *J. exp. Psychol.*, 1959, **57**, 121–129.

Manuscript received 4/1/58

Revised manuscript received 8/22/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

A MODEL FOR ORDERED METRIC SCALING BY COMPARISON OF INTERVALS*

ROBERT F. FAGOT†

UNIVERSITY OF OREGON

This paper presents a model of individual choice behavior for application to experimental situations in which a subject is required to compare utility intervals (differences in subjective value). This model is contrasted with a weaker model, which is also derived. Both models generate ordered metric scales, but differ in predictive power. An experiment on the utility of grades, which provides a test and comparison of the models, is presented.

Coombs [2] introduced the term ordered metric to denote those scales which provide an ordering of the set of alternatives, or stimuli (the single property of an ordinal scale), and in addition at least a partial ordering of the distances between stimuli (utility intervals). Siegel [8] presented a method, based on a one-person game, which generated a so-called higher-ordered metric scale, i.e., a complete ordering on utility intervals. The present model also generates a higher-ordered metric scale. The term ordered metric scale, as used in this paper, will always refer to a higher-ordered metric scale, as the term is used by Siegel [8], and will be abbreviated OM.

Coombs presented no formal characterization of the OM scale, nor did Siegel; one of the purposes of this paper is to present such a characterization. Two important results will be presented in relation to the properties of such scales. First, it will be shown that it is possible to construct a scale which satisfies the above two properties (an ordering of the set of stimuli, and an ordering of utility intervals), but which is inadmissible as an OM scale. This implies, of course, that the intuitive definition of an OM scale given above is inadequate, since it admits to the class of OM scales types which were not originally intended. Second, it will be shown that it is possible to have two or more models, all of which generate OM scales, but which differ in predictive power.

The subject will be required to make judgments in relation to pairs, triads, and tetrads of stimuli. For pairs of stimuli, he will have the traditional task of choosing one member of the pair on some specified basis. For triads

---

*This research was supported in part by Group Psychology Branch, Office of Naval Research, under Contract Nonr 225(17), with Stanford University, and in part by U. S. Public Health Service grant M-2046.

†The author wishes to express his indebtedness to E. W. Adams, University of California, whose constructive criticism did much to improve the quality of this work.

and tetrads of stimuli, he will be required to make comparative judgments of intervals. As an illustration, if three stimuli have been ordered by the method of pairs xyz, then the subject could be required to make the judgment whether y is more similar to x than to z. The particular relation utilized (in this illustration more similar) will depend on the experimental situation, but in all cases the subject will be required to make a judgment which permits an inference as to the ordering of two utility intervals. For example, if as above the stimuli are ordered xyz, and y is judged more like x than like z, then it will be inferred that y is closer on the subject's utility scale to x than to z; i.e., the utility interval between y and z is greater than the utility interval between x and y. If tetrads are used, the judgment will be of the form "the difference (along some specified dimension) between x and y is greater than the difference between z and w."

Although the fundamental relation between pairs of stimuli is interpreted in this paper in terms of direct comparisons of utility intervals, as illustrated above, the formal structure of the model is applicable to any method, such as Siegel's one-person game [8], for which each judgment implies an ordering of two utility intervals.

Formal Structure of the Model

The model is based on three primitive notions. The first is a set K, interpreted as the set of alternatives, or stimuli. The second is a quaternary relation Q on the cartesian product $K \times K$. The relation Q is interpreted such that $xyQzw$ holds when the difference in subjective value (utility) between x and y is judged *greater than* the difference between z and w. If the alternatives are objects among which the subject is stating preferences, then $xyQzw$ implies that the difference in strength of preference (value) between x and y is greater than the difference between z and w; the subject more strongly prefers x to y than z to w. If the four alternatives are tones which differ only in frequency, then $xyQzw$ implies that the difference in pitch between x and y is judged by the subject to be greater than the difference in pitch between z and w. The third primitive is a binary relation P on the set of alternatives K. P is the relation of *strict preference*; xPy holds if and only if the subject strictly prefers x to y.

A numerical interpretation of the relations P and Q is given with the following definition of a quaternary utility function.

DEFINITION 1. Let U be a real-valued function defined over K; then U is a *quaternary utility function* if it satisfies the following conditions for every x, y, z, and w in K:

(a) xPy if and only if $U(x) > U(y)$;

(b) $xyQzw$ if and only if $U(x) - U(y) > U(z) - U(w)$.

The basic assumption of the model may now be stated.

QUATERNARY HYPOTHESIS (QH). There exists a quaternary utility function for the relations P and Q and set K, with the following restrictions.

(a) For every x and y in K, if $x \neq y$, then $U(x) \neq U(y)$.

(b) For every x, y, z, and w in K, if $U(x) > U(y)$ and $U(z) > U(w)$ and $x \neq z$ or $y \neq w$, then $U(x) - U(y) \neq U(z) - U(w)$.

Restrictions (a) and (b) are necessary if the relations P and Q are to have the desired ordering properties, which will be clear after consideration of empirical consequences 1 and 3, presented in the following section. Condition (b), of course, rules out the possibility that an interval $U(x) - U(y)$ equals some other interval $U(z) - U(w)$. However, this exclusion is consistent with the intended application of the model to situations of strict preference.

For every choice that a subject makes, we may write a corresponding inequality in utility values (Definition 1). An experiment would produce a set of inequalities for each subject, and a quaternary utility function exists for a particular subject (the model holds exactly) if and only if the resulting set of inequalities is consistent, i.e., has a solution. There are well-known decision procedures for determining whether or not a set of inequalities has a solution [e.g., 6].

Satisfaction of the QH is sufficient to guarantee the existence of an OM scale, but it is not necessary. This follows from the fact that, as will be shown in the next section, it is possible to derive from the QH consequences which, if satisfied, guarantee the existence of an OM scale, yet these consequences do not exhaust the empirical significance of the QH. This means that the QH may be disconfirmed even if an OM scale exists. Hence one important problem is to find the weakest conditions imposed by the QH which are capable of generating the desired OM scale. An allied problem is to define progressively stronger alternative models in terms of the empirical consequences of the basic hypothesis, so that it will be possible to state precisely how strong a model is sustained by the data. In the following section some of the empirical consequences which follow from the basic hypothesis are specified and in the process two alternative models are defined.

Some Empirical Consequences of the Quaternary Hypothesis

By empirical or observable consequence of the QH is meant any consequence implied by the QH about the P and Q relations themselves. Unfortunately, a result obtained by Scott and Suppes ([7], pp. 16–27) implies that there is no simple set of consequences of the type to be considered which completely exhausts the empirical significance of the QH for all finite sets of stimuli. Derivation of the empirical consequences, and proofs of certain propositions, such as the independence of these consequences, will not be presented, since they are mostly simple, tedious, and mathematically uninteresting.

The first consequence of the QH is that the binary relation P is a *strict simple ordering* of K. The terminology follows that of Suppes [9].

Empirical Consequence 1 (C1). The binary relation P is a *strict simple ordering* of K; P satisfies the following conditions for every x, y, and z in K:

(a) exactly one of the following holds: $x = y$, xPy, yPx;

(b) if xPy and yPz then xPz.

Thus if C1 is satisfied it becomes possible to derive a ranking without ties of the alternatives in K.

One is perhaps used to thinking of this consequence as the sole axiom of an ordinal model, but certain Q relations are predictable from P relations, and hence the ordinal model must be defined to include an axiom formulating the connection between P and Q where it is possible to predict Q preferences from P preferences. For example, if three alternatives are ordered $xPyPz$, then this knowledge alone implies that the distance between x and z is greater than the distance between x and y, and also greater than the distance between y and z. This notion is formalized below.

Empirical Consequence 2 (C2). For every x, y, z, and w in K,

(a) if xPy and yPz, then $xzQxy$ and $xzQyz$;

(b) if xPy, yPz, and zPw, then $xwQyz$.

In most preference experiments, one would hardly expect C2 to be disconfirmed if C1 is satisfied, but this confidence might not be at all sustained in certain psychophysical experiments. In any event, C1–2 are the axioms of the *Ordinal Model*.

One further consequence of the basic hypothesis permits the formulation of a model which generates an OM scale. Consider a subset P of $K \times K$ such that for all $\langle x, y\rangle$ in $K \times K$, $\langle x, y\rangle$ belongs to P if and only if xPy. P then is the subset of $K \times K$ which is the set of ordered pairs xPy. Then one obvious consequence of the QH is that Q is a strict simple ordering of P.

Empirical Consequence 3 (C3). The quaternary relation Q is a *strict simple ordering* of P (Q is asymmetric, transitive, and connected in P); i.e., for every $\langle x, y\rangle$, $\langle z, w\rangle$, and $\langle u, v\rangle$ in P,

(a) exactly one of the following holds: $\langle x, y\rangle = \langle z, w\rangle$, $xyQzw$, or $zwQxy$;

(b) if $xyQzw$ and $zwQuv$, then $xyQuv$.

The intuitive statement made in the introduction that a complete ordering on utility intervals is required for an OM scale corresponds to the formal requirement that Q be a strict simple ordering of P. The restriction to P is a restriction to positive differences in utility.

The fact that C2 and C3 are independent has some interesting implications. It will be recalled that the intuitive definition of an OM scale requires only that C1 and C3 be satisfied. However, C2 could be disconfirmed even if C1 and C3 were satisfied. Hence C1 and C3 are not sufficient for the derivation of an OM scale; C2 is necessary. This is the point of the remark made in the introduction that the intuitive definition of an OM scale as requiring an ordering of the stimuli and an ordering of utility intervals was inadequate.

C1–3 do not exhaust the empirical significance of the QH. This fact, coupled with the fact that these consequences are sufficient to generate an OM scale, leads to the important result that an OM scale may exist even when the QH is disconfirmed. Such a model is called a *weak* ordered metric model to distinguish it from the stronger model which assumes the existence of a quaternary utility function. Empirical consequences 1, 2, and 3 define the *Weak Ordered Metric Model* (hereafter abbreviated as WOM); i.e., C1–3 are the axioms of the WOM. It should be noted that although the term weak is used, the model still generates a higher-ordered metric scale.

Although C1–3 are the set of necessary and sufficient conditions for the existence of an OM scale, it is of considerable interest to formulate further consequences of the QH, for a number of reasons. First, completely separate from the problem of measurement, these consequences could be used for the prediction of choice behavior. Second, if the QH is disconfirmed, then it becomes possible to state precisely in which way it was disconfirmed by examining each of the consequences. Third, the formulation of such consequences permits the construction of progressively stronger models of choice behavior which can be compared in experimental situations. Finally, the two consequences to be presented exhaust the empirical significance of the QH under certain special conditions to be discussed after their presentation. The first of these consequences is presented below.

EMPIRICAL CONSEQUENCE 4 (C4). For all x, y, z, and w in K, if $xyQzw$ then $xzQyw$, $wyQzx$, and $wzQyx$.

This consequence is stated in its most general form, but in the usual experimental situation the observation $xyQzw$ implies xPy and zPw, in which case the consequence could be simplified to "if $xyQzw$ then $xzQyw$" for direct prediction purposes. In any event, for any given ordering of stimuli, C4 will have only one conclusion with empirical content, not three. Its empirical content corresponds to the arithmetical fact that $a > b$ if and only if $a + x > b + x$. For example, suppose four stimuli are ordered $xPyPzPw$. Corresponding to "if $xyQzw$, then $xzQyw$" is, writing capital letters for the utilities of the stimuli, "if $X - Y > Z - W$, then $X - Z > Y - W$." The conclusion can be written $(X - Y) + (Y - Z) > (Y - Z) + (Z - W)$, from which it can be seen that the interval $(Y - Z)$ is common to both sides of the inequality, which can be reduced to $X - Y > Z - W$, which is of course, the premise.

The following example shows that C4 has implications which C1–3 do not. Suppose again that four stimuli are ordered $xPyPzPw$, and further that the six utility intervals are ordered $xwQxzQywQzwQxyQyz$. P and Q are transitive and hence C1 and C3 are satisfied. Inspection of the ordering of the intervals indicates that C2 is also satisfied. Note, however, that the observation $xzQyw$ implies $xyQzw$ according to C4. But $zwQxy$ was observed; therefore C4 is disconfirmed. This means that C4 has implications which C1–3 do not. This result confirms the important conclusion already stated, that C1–3 do not exhaust the empirical significance of the QH, in spite of the fact that these consequences are sufficient to generate an OM scale.

One final consequence of the QH is now presented.

EMPIRICAL CONSEQUENCE 5 (C5). For all x, y, z, w, u, and v in K, if $xyQzw$ and $uxQwv$, then $uyQzv$.

The empirical content of this consequence is readily seen if the relations are written in terms of the corresponding utility values: if $X - Y > Z - W$ and $U - X > W - V$, then $U - Y > Z - V$. Adding the two inequalities of the premise, $(U - X) + (X - Y) > (Z - W) + (W - V)$, which of course simplifies to $U - Y > Z - V$, the conclusion. The arithmetical interpretation is (i) if $m > n$ and $r > s$, then $m + r > n + s$. In a similar way one can show that C5 also has the arithmetical interpretation (ii) if $r > n$ and $m + n > r + s$, then $m > s$.

When an implication of C5 involves four stimuli (e.g., $abQcd$), then this Q relation can also be predicted by C4, although the two predictions may differ. Thus one of the subjects in the experiment to be reported made the following choices: $dfQab$, $bfQad$, $cdQbc$, and $acQcf$. The first choice implies the second by C4, but the third and fourth imply $adQbf$ by C5, which contradicts the second. This shows that C5 has implications which C1–4 do not, since the total set of observations was consistent with C1–4, but C5 was disconfirmed.

If a C5 implication involves only three stimuli (e.g., $xyQyz$), then it cannot be predicted by C4. It also follows that in such a case the premises of C5 must involve only *five* stimuli, not six. Therefore the number of predictions which can be made by C5 but *not* by C1–4 cannot exceed $\binom{n}{5}$, where n is the number of stimuli in K. However, even if a C5-implication is not predictable by C4, it may be predictable by C3. Since there are $\binom{n}{2}$ ordered utility intervals, there will be exactly $\binom{n}{2} - 1$ Q relations which are not predictable by C3. Therefore the *maximum* number of predictions which can be made by C5 but not by C1–4 is $\binom{n}{5}$ or $\binom{n}{2} - 1$, whichever is smaller. This upper bound reflects the maximum additional predictive power of a model which contains C5 as an axiom as well as C1–4.

Although C1–5 do not exhaust the empirical significance of the QH for all finite sets of stimuli, a weaker statement on sufficiency conditions can be made: if restriction is made to predictions based on not more than *two* choices (i.e., of the form "if A then B" or "if A and B then C"), then C1–5 do in fact exhaust the empirical significance of the QH. This means that under such a restriction, any prediction which can be made with the QH can also be made with C1–5. The increase in predictive power resulting from the utilization of more than two choices is slight for small sets of stimuli (less than ten), and hence this is an important result.

In formulating the WOM, the set of necessary and sufficient conditions for the existence of an OM scale have been specified. Such a model may be compared in experimental situations with the *Strong Ordered Metric Model* (hereafter abbreviated SOM), which assumes the existence of a quaternary utility function. Such a strong model is perfectly satisfied if and only if the set of inequalities in utility values corresponding to the set of observed P and Q relations has a solution. However, as pointed out, even if the solution does not exist, the WOM may be satisfied, and hence an OM scale may exist.

Relation to Two Other Theories

The present model will be discussed in relation to two other theories. The first is that used by Siegel [8] to obtain an OM scale, and the second is a model of riskless choice developed by Adams and Fagot [1].

The essential device underlying Siegel's approach is a one-person game in which the subject chooses between two options, each of which is a probability combination of two outcomes. A chance event E, which has an experimentally determined subjective probability of one-half, is defined. The subject is required to choose between two options (x, y) and (z, w). If he chooses the first and E occurs he gets x, while if E does not occur he gets y. The second option is determined in a similar manner.

If one now introduces a quaternary relation R such that $xyRzw$ holds if and only if the subject chooses the option (x, y) when presented with a choice between (x, y) and (z, w), then condition (b), Definition 1, of a quaternary utility function may be modified as follows:

$$\text{(b}'\text{) } xyRzw \text{ if and only if } U'(x) + U'(y) > U'(z) + U'(w).$$

Condition (b′) is based on the assumption that an individual makes choices among alternatives involving risk as if he were trying to maximize expected utility ([8], pp. 212–213).

Thus there are two methods for deriving OM scales, and it would be of considerable interest to determine if the same scales would be derived by both methods. The following hypothesis postulates a relationship between R and Q.

RQ HYPOTHESIS. For every x, y, z, and w in K, $xyRzw$ if and only if $xwQzy$.

This hypothesis is equivalent to the assumption that the utility functions U and U' are the same. Given a set of axioms on Q (such as C1–5), these axioms can be immediately transformed to a set of axioms on R, by means of the RQ hypothesis. It then becomes possible to determine if individual axioms hold for both Q and R, rather than simply testing the stronger RQ hypothesis. It should be noted that the existence of U' is not necessary for the derivation of an OM scale; satisfaction of axioms equivalent to C1–3 is sufficient.

There are several interesting possibilities which could arise from experimental comparisons of the two methods. For example, both U and U' may exist, yet not be the same. Or, neither U or U' may exist, yet an OM scale may exist for both (i.e., the WOM may be satisfied for both Q and R observations), and this scale may not be the same. Finally, one method may produce an OM scale, but not the other.

Another theory, formally quite similar to the present model, is a theory of riskless choice developed by Adams and Fagot [1]. This theory is concerned with subjects' choices between alternatives which are multidimensional, although the theory has been worked out only for the two-dimensional (or two-component) case. Such alternatives might be political candidates who are described as varying in two characteristics, liberality and foreign policy, for example. In this case the subject would be required to state a preference for one of every pair of such candidates. Or, in an extension, the alternatives may be pairs of objects of any sort. For example, one two-component alternative could be book x and book y, and the subject might have to choose between this alternative and a second which consisted of two books z and w. It is clear then that the components of the two-component alternatives can be thought of as belonging to a set K of alternatives, such as has been considered previously in this paper. The model assumes that the individual behaves as if he assigns subjective values (utilities) to each of the components *independently*, and then *adds* the values together to get the value of the composite alternative. The fundamental assumption of the model is the hypothesis of the existence of this *additive* utility function. Comparable to condition (b), Definition 1,

(b″) $(x, y)P(z, w)$ if and only if $U''(x) + U''(y) > U''(z) + U''(w)$,
where P is the relation of strict preference.

The model generates an OM scale, and a simple transformation relates this model to the present model. This relationship is indicated by the following hypothesis.

PQ HYPOTHESIS. For every x, y, z, and w in k, $(x, y)P(z, w)$ if and only if $xzQwy$.

This hypothesis is equivalent to the assumption that the utility functions U and U'' are the same. All three models are related by the following corollary.

Corollary of RQ and PQ Hypotheses. For every x, y, z, and w in K, $(x, y)P(z, w)$ if and only if $xyRzw$.

It should be realized that R holds between options and P holds between alternatives. Thus in the case of P the outcome is x *and* y, whereas in the case of R the outcome is x *or* y, depending upon the outcome of a chance event E.

In this additive model of riskless choice, the basic assumption of the additivity of the components is usually used simultaneously to determine the utility values and to make predictions. Under these conditions, the predictive consequences of the model are much weaker than those using utility values obtained outside the model by independent methods of measurement. Observations on Q or R relations could provide this independent method of measurement. In other words, U or U' could be used to assign utility values to the alternatives in a basic set K, and then the additive model could be used to make predictions between *pairs* of these alternatives, under the assumption that the utility values are additive.

Some Experimental Results

Some results of an experiment on the utility of course grades afford a test of the model and illustrate the derivation of an OM scale. Ten students in introductory psychology served as subjects. The alternatives in the set K are the course grades A, B, C, D, and F. Appropriate operational definitions were given to the relations P and Q such that aPb means that the subject prefers the grade A to the grade B, and $dfQab$ means that the difference, in value (utility), between the grades D and F is greater than the difference between the grades A and B. Care was taken to insure that each subject thoroughly understood the operation of comparing two utility intervals and reporting the interval which was subjectively greater.

Let n be the number of stimuli. Then N_0, the number of Q relations predictable from a knowledge of an ordering of the stimuli, is

$$N_0 = 2\binom{n}{3} + \binom{n}{4}. \tag{1}$$

In this experiment there are five stimuli in K; if all comparisons are made, the complete data for each subject would consist of 10 P relations and 45 Q relations. Of these 45 observations, 25 would be predictable from C2 ($N_0 = 25$). However, in a preference experiment of this sort, there seems little necessity for making a complete test of this consequence, so only 5 such comparisons were presented to each subject. The responses of each of the ten subjects were consistent with C2. C1 was not tested; it was assumed that all subjects preferred the grades in the order $aPbPcPdPf$.

The choices of one of the subjects will be selected for discussion. Table 1 lists the 20 choices of this subject which cannot be predicted from a knowledge of an ordering of the stimuli, and hence these observations contain the ordered metric information.

These choices are divided into subsets by means of C4; subset B is predictable from subset A by means of C4. Utilizing this consequence one can predict (16) from (2), (17) from (5), (18) from (3), (19) from (12) and (20) from (8); thus C4 is perfectly satisfied.

TABLE 1

Observations of One Subject in Experiment on Utility of Grades

Subset A			Subset B
1. abQbc	6. bfQab	11. acQcd	16. acQbd
2. abQcd	7. bcQcd	12. dfQac	17. bfQac
3. dfQab	8. dfQbc	13. cfQac	18. bfQad
4. bdQab	9. cfQbc	14. adQdf	19. cfQad
5. cfQab	10. dfQcd	15. dfQbd	20. cfQbd

A necessary and sufficient condition for admission of an observation to subset A is that the two intervals corresponding to the two ordered pairs of a Q relation do not overlap. For example, if $xPyPz$, then the only such relation is $xyQyz$ (or $yzQxy$, whichever is observed). The number of Q relations in subset A is

$$N_A = \binom{n+1}{4}. \tag{2}$$

The two intervals for a Q relation in subset B always overlap. For example, observation (16) implies $A - C > B - D$, and it is clear that these two intervals have the interval $B - C$ in common. The number of observations in subset B, and hence the number of predictions that can be made in any experiment by means of C4, provided that certain observations are made in subset A, is

$$N_B = \binom{n}{4}. \tag{3}$$

It is important to note that although N_B choices can be predicted from certain choices in subset A, these predictions cannot be made by the WOM, since C4 is not an axiom in this model.

Inspection of Table 1 shows that the choices of this subject were consistent with C3 (Q is transitive), and with C4. He satisfies not only the WOM but also the SOM (the set of inequalities corresponding to the observations have a solution). The OM scale which is derived from the observations in Table 1 is defined by the following ordering of utility intervals

$$A - F > B - F > C - F > A - D > D - F > A - C > B - D > A - B > B - C > C - D. \tag{4}$$

In general, if there are i intervals, then $(i - 1)$ Q relations are sufficient to define the OM scale—but not always necessary. For example, of the nine Q relations corresponding to (4), the following seven are sufficient to generate the OM scale by the WOM: *cfQad*, *adQdf*, *dfQac*, *acQbd*, *bdQab*, *abQbc*, *bcQcd*. The relations *afQbf* and *bfQcf* are not necessary, since they are implied by the P relations by means of C2. Thus the remaining 38 Q relations contain superfluous information since they are predictable from these seven by means of the axioms of the WOM, thus providing a test of the model. Depending on the structure of the scale, as few as three observations may in some cases provide sufficient information for the construction of the OM scale. Unfortunately, precisely which observations provide this information cannot be specified in advance. However, Siegel ([4], p. 141) has developed a "maximin" method which maximizes the number of predictions which may be made from observing a minimum number of choices, and specifies the order in which choices should be examined in constructing the scale. His method may be adapted for use with the present model.

The number of observations necessary and sufficient to represent the entire set of observations can be taken as a measure of the *predictive power* of the model. The smaller the number of such observations, the greater is the predictive power of the model. In the case above, two of the observations, *cfQad* and *acQbd* are not necessary in the SOM since they are implied by *dfQac* and *abQcd*, respectively, by C4. Thus in the SOM only five observations are necessary, whereas in the WOM seven are necessary, indicating the correspondingly greater predictive power of the former. In this experiment the number of choices necessary and sufficient to represent the total set of 45 observations ranged, in the WOM, from a low of six for one subject to a high of eight for another subject, and in the SOM, from a low of three to a high of five, the lower values of the SOM being indicative of its greater predictive power. This concept of predictive power is treated more extensively in other sources ([5]; [3], pp. 28–36).

The general results of the experiment are as follows.

1. Seven of the ten subjects perfectly satisfied the WOM; hence for these seven subjects it was possible to construct an OM scale.

2. Two of these seven failed to satisfy the SOM. These two subjects are of special interest, since for them it was possible to construct an OM scale in spite of the fact that they did not satisfy the QH. One of these two had a

single disconfirmation of C4, and the second had a single disconfirmation of C5.

3. Seven of the ten subjects perfectly satisfied C4 (five times each); the remaining three each had a single disconfirmation.

4. At most *one* prediction for each subject can be made by C5 which cannot be made by means of C1–4, since there are only five stimuli in this experiment. The results were that one such prediction was made for each of three subjects, and one of these predictions was disconfirmed. (If the number of stimuli were increased by one to six, then the maximum number of implications of C5 for each subject would increase from one to six.)

5. For each of the three subjects who did not perfectly satisfy the WOM, the reversal of a single choice would have resulted in satisfaction of the model; however, one of these three still would not have satisfied the SOM. Methods of analyzing the fit of the model have been devised, but discussion will be deferred to a later paper on the analysis of a series of such experiments.

REFERENCES

[1] Adams, E. W. and Fagot, R. F. A. model of riskless choice. *Behav. Sci.*, 1959, **4**, 1–10.

[2] Coombs, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*, 1950, **57**, 145–158.

[3] Fagot, R. F. An ordered metric model of individual choice behavior. Tech. Rep. No. 13, Contract Nonr 225(17), Stanford Univ., 1957.

[4] Hurst, P. M. and Siegel, S. Prediction of decisions from a higher ordered metric scale of utility. *J. exp. Psychol.*, 1956, **52**, 138–144.

[5] Kemeny, J. and Oppenheim, P. Systematic power. *Phil. Sci.*, 1955, **22**, 27–33.

[6] Kuhn, H. W. Solvability and consistency for systems of linear equations and inequalities. *Amer. math. Mon.*, 1956, **63**, 217–232.

[7] Scott, D. and Suppes, P. Foundational aspects of theories of measurement. Tech. Rep. No. 6, Contract Nonr 225(17), Stanford Univ., 1957.

[8] Siegel, S. A method for obtaining an ordered metric scale. *Psychometrika*, 1956, **21**, 207–216.

[9] Suppes, P. *Introduction to logic.* Princeton: Van Nostrand, 1957.

Manuscript received 9/29/58

Revised manuscript received 10/24/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

A NOTE ON FACTOR ANALYSIS: ARBITRARY ORTHOGONAL TRANSFORMATIONS

EDWARD E. CURETON

UNIVERSITY OF TENNESSEE

A modification of the Gram-Schmidt process yields an easily constructed orthogonal transformation matrix which may be used to rotate a centroid, principal axis, or maximum likelihood factor matrix in a manner such that one of the new axes has predetermined direction. The procedure is illustrated by rotating a centroid factor matrix into an abbreviated bifactor matrix, the general factor being defined as the centroid of a specified subgroup of reasoning tests.

Starting with an n by m orthogonal factor matrix A, such as a centroid, principal axis, or maximum likelihood matrix, with entries a_{ik} representing the loadings of n tests or other initial variables on m factors, it is decided on substantive (rather than mathematical) grounds that one rotated axis should pass through a particular point. The other axes may be anywhere, so long as they remain orthogonal to the one chosen. The problem is to construct an orthogonal transformation matrix to accomplish this purpose.

Reyburn and Taylor have discussed this problem [2], and have presented a method for constructing the desired transformation matrix. Their method, however, is fairly complicated, though less so, as they point out, than the direct application of the Gram-Schmidt process. By taking advantage of the fact that all of the axes but one may be defined arbitrarily, the Gram-Schmidt process may be simplified to the point where the desired transformation matrix can be computed with much less effort than is required by the Reyburn-Taylor procedure.

The general Gram-Schmidt process converts any matrix X into an orthogonal matrix Z whose first row is a unit vector collinear with the first row vector of X. If the first row vector of X is a unit vector, the first row vector of Z is identical to it. In this case the general Gram-Schmidt equations are

$$
\begin{aligned}
Z_1 &= X_1 \,, \\
Z_2 &= [X_2 - (Z_1 \cdot X_2) Z_1]/K_2 \,, \\
Z_3 &= [X_3 - (Z_1 \cdot X_3) Z_1 - (Z_2 \cdot X_3) Z_2]/K_3 \,, \\
Z_4 &= [X_4 - (Z_1 \cdot X_4) Z_1 - (Z_2 \cdot X_4) Z_2 - (Z_3 \cdot X_4) Z_4]/K_4 \,, \\
Z_5 &= [X_5 - (Z_1 \cdot X_5) Z_1 - (Z_2 \cdot X_5) Z_2 - (Z_3 \cdot X_5) Z_3 - (Z_4 \cdot X_5) Z_4]/K_5 \,,
\end{aligned}
$$

etc. The Z_i and X_i are the row vectors of Z and X, the $Z_i \cdot X_j$ in parentheses are scalar products, and the K_i are normalizing factors.

Let the selected unit vector be $a, b, c, d, e, \cdots$, and let X be of the form,

$$X = \begin{bmatrix} a & b & c & d & e \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} .$$

The scalar products of the Gram-Schmidt equations then become single elements of Z, and

$$\begin{aligned} Z_1 &= X_1 , \\ Z_2 &= [X_2 - aZ_1]/K_2 , \\ Z_3 &= [X_3 - bZ_1 - Z_{22}Z_2]/K_3 , \\ Z_4 &= [X_4 - cZ_1 - Z_{23}Z_2 - Z_{33}Z_3]/K_4 , \\ Z_5 &= [X_5 - d\, Z_1 - Z_{24}Z_2 - Z_{34}Z_3 - Z_{44}Z_4]/K_5 , \end{aligned}$$

etc. These equations can be solved literally for the two-variable, three-variable, four-variable, and five-variable cases. First define

$$\begin{aligned} k_2^2 &= 1 - a^2, \\ k_3^2 &= 1 - a^2 - b^2, \\ k_4^2 &= 1 - a^2 - b^2 - c^2. \end{aligned}$$

These k's are related but not identical to the K's of the previous equations. The desired transformation matrices are then

$$Z = \begin{bmatrix} a & b \\ b & -a \end{bmatrix} ,$$

$$Z = \begin{bmatrix} a & b & c \\ k_2 & -ab/k_2 & -ac/k_2 \\ 0 & c/k_2 & -b/k_2 \end{bmatrix} ,$$

$$Z = \begin{bmatrix} a & b & c & d \\ k_2 & -ab/k_2 & -ac/k_2 & -a\, d/k_2 \\ 0 & k_3/k_2 & -bc/k_2k_3 & -b\, d/k_2k_3 \\ 0 & 0 & d/k_3 & -c/k_3 \end{bmatrix} ,$$

$$Z = \begin{bmatrix} a & b & c & d & e \\ k_2 & -ab/k_2 & -ac/k_2 & -a\,d/k_2 & -ae/k_2 \\ 0 & k_3/k_2 & -bc/k_2k_3 & -b\,d/k_2k_3 & -be/k_2k_3 \\ 0 & 0 & k_4/k_3 & -c\,d/k_3k_4 & -ce/k_3k_4 \\ 0 & 0 & 0 & e/k_4 & -d/k_4 \end{bmatrix}.$$

The system by which larger matrices may be constructed will be clearer if it is noted that $k_1 = 1$ and that the final k of each matrix (not shown as such in the last row of each of those above) is equal to the last element of the first row. Thus for the 5 by 5 matrix, $k_5^2 = 1 - a^2 - b^2 - c^2 - d^2 = e^2$, and $k_5 = e$, since Z_1 is a unit vector. With this notation,

$$Z = \begin{bmatrix} a & b & c & d & e \\ k_2/k_1 & -ab/k_1k_2 & -ac/k_1k_2 & -a\,d/k_1k_2 & -ae/k_1k_2 \\ 0 & k_3/k_2 & -bc/k_2k_3 & -b\,d/k_2k_3 & -be/k_2k_3 \\ 0 & 0 & k_4/k_3 & -c\,d/k_3k_4 & -ce/k_3k_4 \\ 0 & 0 & 0 & k_5/k_4 & -d\,e/k_4k_5 \end{bmatrix},$$

and for the two-, three-, or four-variable case only the first two, three, or four rows and columns are used.* The required transformation is then

$$AZ' = G,$$

G being the rotated factor matrix.

Example

The A matrix of Table 1 is a centroid matrix computed by the writer from a table of intercorrelations reported by Swineford ([3], p. 14, Table 4). Her subjects were 504 school children of both sexes in grades 5 through 12, all between the ages of 12–0 and 14–5, and all from the Chicago metropolitan area. The centroid solution was iterated ten times to obtain communalities all stable within ± .001.

Swineford argues on substantive (psychological) grounds that a general factor can be defined by the first three tests, which appear to the present writer to be reasoning tests. Accepting her argument, the centroid matrix will be rotated orthogonally so that one axis passes through the centroid of tests 1, 2, and 3. The first row of the transformation matrix Z consists of the normalized sums of the first three entries in each column of A. The rest of the transformation matrix is constructed by the Gram-Schmidt formulas for the three-variable case. Thus

---

*I am indebted to my colleague, Dr. Frederick A. Ficken, Department of Mathematics, University of Tennessee, for a proof that Z, as defined by this last matrix, is actually orthogonal.

$$Z = \begin{bmatrix} .9059 & .2405 & -.3485 \\ .4235 & -.5144 & .7455 \\ 0 & -.8229 & -.5679 \end{bmatrix}.$$

TABLE 1

Factor Matrices (Swineford Data; N= 504)

Name of test	A				G			F			S			
	a_{i1}	a_{i2}	a_{i3}	h^2	g_{i1}	g_{i2}	g_{i3}	f_{i1}	f_{i2}	f_{i3}	s_{i1}	s_{i2}	s_{i3}	h^2
1.Arith.Comput.	659	285	-308	610	773	-097	-060	773	-111	-026	677	000	000	458
2.Series Complet.	643	065	-243	477	683	058	085	683	101	-019	688	000	000	473
3.Deduction	619	160	-188	444	665	040	-025	665	011	046	661	000	000	437
4.General Inf.	802	228	235	750	699	398	-321	699	054	508	757	000	415	745
5.Reading Comp.	730	202	106	585	673	284	-226	673	041	361	718	000	265	586
6.Word Meaning	756	261	397	797	609	482	-440	609	030	652	668	000	622	833
7.Punched Holes	716	-512	-085	782	555	503	470	555	688	023	578	652	000	759
8.Drawings	660	-420	029	613	487	517	329	487	598	133	537	577	000	621
9.Vis.Imagery	638	-266	057	481	494	450	187	494	450	. 186	554	413	000	477
Sum	6223	003	000	5539	5638	2635	-001	5638	1862	1864				5389
Check					5638	2634	-002	5638	1863	1864				

Note: All entries have three decimal places.

Postmultiplying A by Z' is equivalent to multiplying A by Z, row by row. Doing this, the entries in matrix G of Table 1 are obtained. Those in the g_{i1} column are the loadings on the general factor defined by the centroid of tests 1, 2, and 3. These tests have small loadings summing to zero within rounding error in columns g_{i2} and g_{i3}, as they should. For the rest, columns g_{i2} and g_{i3} represent a residual centroid matrix; all entries in column g_{i2} except those of the first three tests are positive and substantial, and those in column g_{i3} sum to zero within rounding error. If this were not already a residual centroid matrix, it could be made into one by normalizing the sums of the last six entries in columns g_{i2} and g_{i3}, using these values as the first row of a new two-variable Z, completing this Z by the Gram-Schmidt formulas, and postmultiplying columns g_{i2} and g_{i3} by Z'.

To complete the derived bifactor solution, the residual centroid should be located at the center of the positive quadrant. Hence the axes of g_{i2} and g_{i3} should be rotated so that the new axes will be at 45 degree angles to the first residual centroid axis. To do this one employs the two-variable Landahl transformation [1],

$$L = \begin{bmatrix} .7071 & .7071 \\ .7071 & -.7071 \end{bmatrix}.$$

Postmultiplying columns g_{i2} and g_{i3} of G by L', yields the columns f_{i2} and f_{i3} of matrix F, Table 1. Column f_{i1} is identical to column g_{i1} of matrix G, and matrix F is the derived bifactor solution. Matrix S is Swineford's original bifactor solution, computed by Holzinger's methods.

Considering first the similarities between the two solutions, it is clear that the first factor is a general factor, the second factor a space factor, and the third factor a verbal factor. So far as gross factorial interpretation is concerned, the two solutions are essentially equivalent.

Considering the differences between the two solutions, it is the writer's opinion that the derived solution is superior on three counts. First, it uses all of the common variance. This property, admittedly, depends upon the fact that the centroid solution was iterated, a tedious procedure at best. But the centroid communalities sum to 5.539, while the direct bifactor communalities sum to only 5.389. The distributions of the third-factor centroid residuals (at the tenth iteration) and the direct bifactor residuals appear below.

Residual:	070 089	050 069	030 049	010 029	−010 009	−030 −011	−050 −031	−070 −051	−090 −071	−110 −091
Centroid:			1	9	18	6	2			
Bifactor:	1	1	7	3	14	5	3		1	1

The centroid method clearly gives a better factorial fit. With larger matrices, moreover, the number of iterations required to stabilize the communalities is much reduced.

The second reason for preferring a derived solution is that in it errors are averaged over whole columns of the correlation matrix, and distributed over *all* factor loadings, both before and after rotation. In the direct bifactor method, near-zeros become exact zeros, and all error variance is forced into the nonzero loadings.

The third reason for preferring a derived solution is that it shows the secondary relationships more clearly. In columns f_{i2} and f_{i3} of F, all entries other than the first three are positive. Hence it is clear at once that there is some variance common to the verbal and space factors over and above what they have in common with the reasoning-general factor. In the direct bifactor solution, the general factor is not wholly defined by the three reasoning tests; it includes also most of the variance common to the space tests and the verbal tests. A certain amount of additional "forcing" is necessary to achieve this result; it cannot be produced by rotations in 3-space without moving the general-factor axis away from the centroid of the three reasoning tests, in which case these tests acquire larger loadings on the other two factors.

This is, in fact, an *abbreviated* bifactor solution. A *complete* bifactor solution ([4], ch. XIX) requires a general factor *in addition* to as many group factors as there are simple-structure factors. The general factor is then defined by the correlations between the primary factors of the oblique simple structure rather than by any one subgroup of the tests, and the general-factor loadings are obtained by projecting the second-order factor loadings back into the first-order domain.

REFERENCES

[1] Landahl, H. D. Centroid orthogonal transformations. *Psychometrika*, 1938, **3**, 219-223.

[2] Reyburn, H. A. and Taylor, J. G. Some factors of personality. *Brit. J. Psychol., Gen. Sec.*, 1939, **30**, 151-165.

[3] Swineford, F. A study in factor analysis: the nature of the general, verbal, and spatial bi-factors. *Supplementary educ. Monogr.*, 1948, No. 67.

[4] Thomson, G. H. *The factorial analysis of human ability.* (3rd ed.) New York: Houghton Mifflin, 1948.

Manuscript received 6/5/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

RANDOMLY PARALLEL TESTS AND LYERLY'S BASIC ASSUMPTION FOR THE KUDER-RICHARDSON FORMULA (21)

Frederic M. Lord

EDUCATIONAL TESTING SERVICE

The K-R (21) formula can be derived without recourse to the undesirable assumption that the test items are all indistinguishable.

Certain of the statements made by Lyerly [4] about the assumptions underlying the K-R (21) formula will mislead readers to reject it on the grounds that the assumptions are rather implausible.* The present note is written in the hope of preventing such a conclusion.

Lyerly ([4], p. 269) states "The basic assumption for K-R (21) ··· implies that individual items have no separate identities, even though each may be worded differently or may ask a question different from that asked by any other item. Furthermore, item equivalence from person to person does not exist. The same item, worded identically, may appear on everyone's test booklet, but item 1 for Subject A is regarded as being no more like item 1 for Subject B than it is like item 2 for Subject B." A few sentences later, he asserts that Lord's derivation [1] "implicitly uses this assumption."

It is a fact that K-R (21) can be derived from the fundamental assumption just quoted. It does not follow that the derivation of K-R (21) in [1], or other derivations, "implicitly use" this assumption, or imply this implication. K-R (21) was originally derived on the assumption that all the test items were equivalent (of equal difficulty and discriminating power); it does *not* follow that the derivation in [1] is based on the same assumption, although this too has been asserted. If two sets of assumptions lead to the same conclusions in all possible applications, then they may be considered, in a sense, identical assumptions; but the fact that they lead to the same conclusion in some particular case does not prove that the assumptions are identical, nor that their implications are the same.

From one point of view, the derivation of K-R (21) in [1] does not rest on "assumptions" at all. The derivation simply asserts the following.

Prepare a very large pool of dichotomous achievement or aptitude items of any kind or kinds whatsoever. Draw randomly and independently from this pool a number of samples of $n \geq 2$ items each. Each sample will be treated as an n-item test, administered (with ample time limit) to a group of examinees, and scored by counting the number of right answers for each examinee. (In actual practice special measures will be needed to minimize fatigue and practice effects, and also other types of variability within the individual examinee; these admittedly are not taken into account by the derivation, but they are irrelevant to the point at issue here.)

---

*While Lyerly's assumptions may seem rather implausible for ordinary mental testing, they are highly appropriate for certain other situations—for example, for rating situations where no two subjects are rated by the same judges. Lyerly's results should be of real value in such situations. The present note deals only with the usual type of objective test, as do the references cited.

Now, take the scores of a particular examinee on the various n-item tests and compute an unbiased estimate of their standard deviation. This is his estimated standard error of measurement. Obtain an average (more precisely, the quadratic mean) of the estimated individual standard errors of measurement. What the derivation in [1] asserts is that the average standard error of measurement so obtained is the same (except for sampling fluctuations) as that obtained from K-R (21) for any of the n-item tests by the usual formula

$$\text{S. E. Meas.} = s_x \sqrt{1 - r_{xx}} \,,$$

where r_{xx} is K-R (21).

The foregoing assertion, made in [1], is experimentally verifiable, providing fatigue, practice effect, and other temporal variability in the examinee can be minimized. If temporal effects are eliminated from consideration, so that the order in which the items are administered is immaterial, then it will make no difference whether the various n-item tests are drawn by random sampling in advance of the testing, or whether the pool of items is administered first and the random samples of n items drawn afterwards. In the latter case, the experiment is identical with that performed in an elementary statistics course of drawing black and white balls from an urn. The results that will be obtained has been well known since the time of Jacob Bernoulli.

When the examinees took an n-item test in the situation just outlined, the n items had their own clear identities. This is manifested, for example, by the fact that in the case of an easy item most (or even all) examinees answer it correctly whenever it is administered. If n is small, one n-item test may be appreciably easier than another simply because it happens to contain several items that are unusually easy. (This means that the errors of measurement for examinee A are *not* independent, over tests, of the errors of measurement for examinee B; but this fact need not detain us here.)

The n-item tests described are called "randomly parallel" tests. In actual practice, "parallel" tests are not constructed by random sampling (although in many cases this would probably be better than reliance on sheer subjective judgment). On the other hand, neither are tests constructed so well that they are strictly parallel according to the assumptions underlying K-R (20), as evidenced by the fact that "parallel" forms usually have slightly different score means, variances, and so forth. Although neither mathematical model describes exactly any actual pair of so-called "parallel" tests, both models are very useful in mental test theory. The utility of the randomly parallel test model is evidenced by some of the practical conclusions to which it has led, as shown in [2] and [3]. It is desirable, therefore, that this model should not be identified with the assumption given in quotes at the beginning of this note.

REFERENCES

[1] Lord, F. M. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 1955, **20**, 1–22.

[2] Lord, F. M. Tests of the same length do have the same standard error of measurement. *Educ. psychol. Measmt*, in press.

[3] Lord, F. M. Statistical inferences about true scores. *Psychometrika*, 1959, **24**, 1–17.

[4] Lyerly, S. B. The Kuder-Richardson formula (21) as a split-half coefficient, and some remarks on its basic assumption. *Psychometrika*, 1958, **23**, 267–270.

Manuscript received 9/10/58

PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

ESTIMATING ITEM INDICES BY NOMOGRAPHS*

ROBERT M. COLVER

DUKE UNIVERSITY

Two nomographs are presented for estimating item validity indices identical in value to those obtained from Flanagan's table and to those obtained from Davis' chart. Experience has shown that the use of the nomographs results in the saving of a significant amount of time with no loss in accuracy. The nomographs also provide a method of quick conversion between the familiar coefficients and the Davis indices, which are less familiar but which offer greater flexibility.

In evaluating the effectiveness of a test, the difficulty and the discriminatory power of the component items have significant bearing. The usability, reliability, and validity of the test are in part determined by these characteristics.

Some of the more common statistical indices currently in use to evaluate test items are: critical ratio, phi coefficient, biserial correlation, point biserial correlation, tetrachoric correlation, and product moment correlation. Although these indices are based on different underlying assumptions, all involve computational procedures that become laborious. The calculation needed for the necessarily substantial number of items in a test often represents more time and effort than are available to many test makers.

Two general methods have been developed to reduce this computational labor: graphic methods (nomographs and Abacs) for estimating the various coefficients, and tabular presentations of coefficients and indices. These tables are based primarily on procedures outlined by Kelley [9].

Nomographs for estimating critical ratios or t's were developed by Appel [1] and Arnold [2]. Dunlap [4] modified the standard Dunlap and Kurtz Nomograph [5] to estimate biserial correlations for test-item analysis purposes. Kuder [10] devised a set of nomographs to estimate point biserial r, biserial r, and four-fold correlations. Lawshe [11] presented a nomograph, based on the Kelley procedures, for estimating the discriminatory values of test responses using a scale of "D-values" of + 5 to − 5.

Guilford [8] constructed an Abac for estimating the significant (and very significant) levels of phi coefficients. Fan [6], in developing an item analysis

---

*© Robert M. Colver, 1959.

table, also used an Abac, as did Flanagan [7]. The use of either of these Abacs for reading item indices would be difficult because of the complexity of their diagrams.

Currently the most widely used techniques are those involving the reading of tables based primarily on Kelley's procedure. Probably the one most satisfactory item validity index is that obtainable from Flanagan's table ([13], pp. 345–351), which provides an estimate of the correlation between responses to an item and total test score. One limitation in the use of Flanagan's estimated coefficient is that units on the scale of correlation values are not linear; that is, the difference between correlations of .25 and .30 is not comparable to a difference between correlations of .93 and .98. Davis [3, 12] developed a table of discrimination indices based on a z transformation of the correlation coefficient multiplied by a constant. This provides indices for which the units are approximately equal throughout the scale, and avoids the use of decimal points. Davis also includes a difficulty index based on a linear transformation of the proportions of success. The Davis table so far has had limited use in spite of the advantages of the linearity of his indices, because as Thorndike [14] suggested "... the numerical values of the proposed new indices will be entirely unfamiliar to the user and will present some difficulty of interpretation on that account."

While the tabular presentations of Flanagan and of Davis have greatly simplified the computational problems involved in determining item validities, both presentations have minor limitations. Flanagan's table, as it is usually published, has a number of pages of tables to be consulted; and Davis' table, although usually on a single page, is of an unwieldy size. Both tables require that the user make four-way interpolations in many of the readings.

The nomographs presented here combine the speed and simplicity of nomograph reading with the familiarity of Flanagan's table and the advantages of Davis' indices: they also provide a method of quick conversion between the familiar Flanagan coefficients and Davis' indices, which are less familiar but which offer greater flexibility and convenience in use.

The Nomographs

The nomographs* shown as Figures 1 and 2 are entered with the same data used to enter either the Davis table or the Flanagan table. From them the following indices may be obtained: a product moment correlation, the Davis discrimination index, an estimated per cent of correct responses in the entire sample, and the Davis difficulty index.

To enter these nomographs the percentage of success on a test item of an upper group and the percentage of success on a lower group are needed.

*The nomographs as presented here are too small for efficient use. Single copies of larger-sized reproduction may be obtained by writing the author.

As Kelley [9] has shown, these groups should be based on the higher 27% of the total sample and the lower 27% of the total sample. (In entering the difficulty nomograph these particular percentages are recommended simply because they are needed to enter the discrimination nomograph.)

The percentage of success of each of the two groups can be determined by calculating the per cent of each group that *marks* the correct answer, but Davis ([12], p. 277) has shown that the number of examinees that actually *knows* the answer is a more fundamental estimation; therefore, it is recommended that the percentages of success of each group be determined by the formula suggested by Davis ([3], p. 6):

$$P = \frac{R - [W/(K - 1)]}{N - NR},$$

where

P = the proportion of the sample that *knows* the answer to an item,
R = the number of testees in a sample that answer the item correctly,
W = the number of testees in the sample that answer the item incorrectly,
N = the number of testees in the sample,
NR = the number of testees in the sample who do not reach the item in the in the time limit,
K = the number of choices in the item.

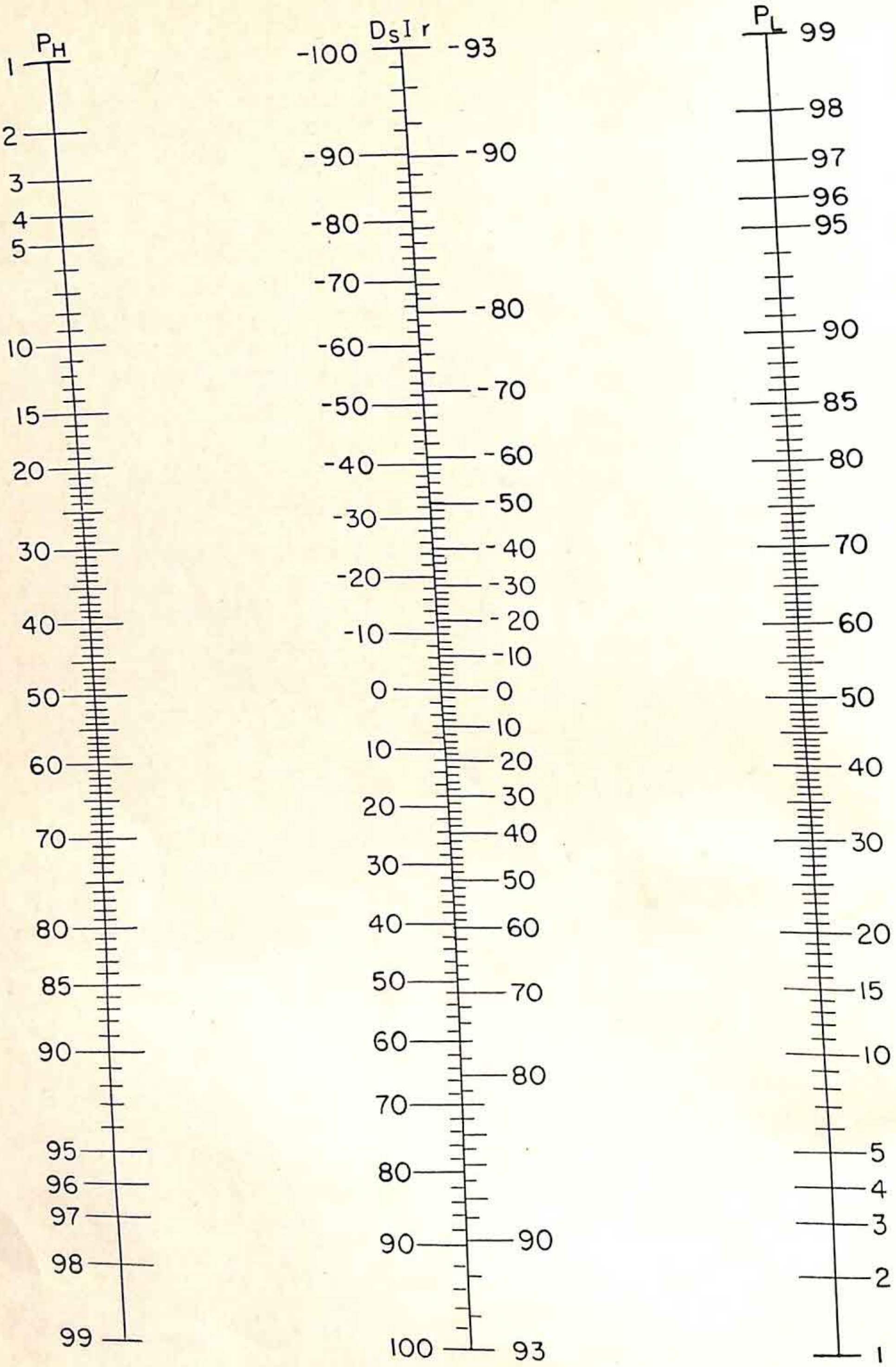
The Discrimination Nomograph (Figure 1)

To obtain the product moment correlation or the item discrimination index locate the determined percentage of success of the higher group on scale P_H and the determined percentage of success of the lower group on scale P_L. Lay a straight edge (a transparent one is preferable) through these points. The discrimination indices are then read at the point where the straight edge crosses the center scale. The right side of the scale, r, gives product moment correlations and the left side of the scale, D_SI, gives the discrimination index.

A product moment correlation, determined by any method, can be directly converted to a discrimination index by locating the r value on the r side of the center scale and reading its corresponding value on the D_SI side of the center scale. Similarly the D_SI values can be converted directly to r values.

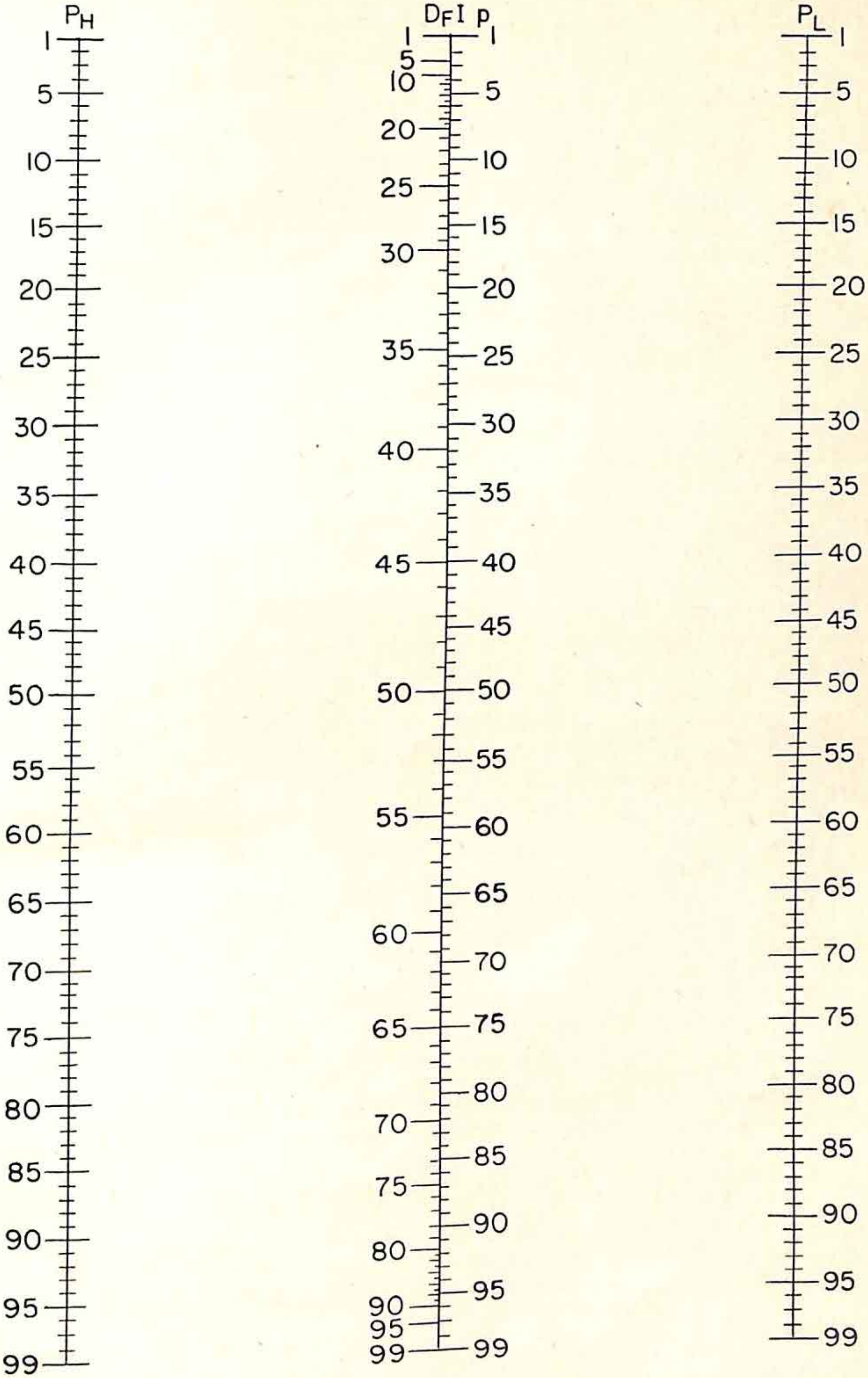
The Difficulty Nomograph (Figure 2)

To obtain an estimate of the percentage of correct responses in the entire sample and the difficulty index locate the determined percentage of success of the higher group on scale P_H and the determined percentage of success of the lower group on scale P_L. Lay a straight edge through these



DISCRIMINATION NOMOGRAPH

FIGURE 1



DIFFICULTY NOMOGRAPH

FIGURE 2

points. The indices are then read at the point where the straight edge crosses the center scale. The right side of the scale, p, gives an estimate of the percentage of correct responses in the entire sample and the left side of the scale, D_FI, gives the difficulty index.

Direct conversion of percentage of success, determined by any method, to the difficulty index can be made by locating the value of p on the p side of the center scale and reading its corresponding difficulty index on the D_FI side of the center scale. Similarly a difficulty index can be converted directly to an estimated percentage of success of an entire sample.

Other Uses and Limitations

It has been suggested ([12], p. 298) that "Flanagan's table has general utility and may be found exceptionally useful in many instances outside the field of test construction where economical approximations of biserial correlations are required." Similar use could be made of the r's obtained from this discrimination nomograph, as the values are identical to those obtained from Flanagan's table.

The literature on the use of Flanagan's and Davis' tables points out some limitations for each table. Similar limitations would apply to the use of these nomographs.

REFERENCES

[1] Appel, V. Companion nomograph for testing the significance of the difference between uncorrelated percentages. *Psychometrika*, 1952, **17**, 325–330.

[2] Arnold, J. N. Nomogram for determining validity of test items. *J. educ. Psychol.*, 1935, **26**, 151–153.

[3] Davis, F. B. *Item-analysis data: their computation, interpretation, and use in test construction.* Cambridge: Harvard Univ. Press, 1949.

[4] Dunlap, J. W. Nomograph for computing biserial correlations. *Psychometrika*, 1936, **1**, 59–60.

[5] Dunlap, J. W. and Kurtz, A. K. *Handbook of statistical nomographs, tables, and formulas.* New York: World Book, 1932.

[6] Fan, C. T. Note on construction of an item analysis table for the high-low-27-percent group method. *Psychometrika*, 1954, **19**, 231–236.

[7] Flanagan, J. C. General consideration in the selection of test items and a short method of estimating the product moment coefficient from the data at the tails of the distributions. *J. educ. Psychol.*, 1939, **30**, 674–680.

[8] Guilford, J. P. The phi coefficient and chi squares as indices of item validity. *Psychometrika*, 1941, **6**, 11–19.

[9] Kelley, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, **30**, 17–24.

[10] Kuder, G. F. Nomograph for point biserial r, biserial r, and four-fold correlations. *Psychometrika*, 1937, **2**, 135–138.

[11] Lawshe, C. H. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, **26**, 846–849.

[12] Lindquist, E. F. (Ed.) *Educational measurements.* Washington: Amer. Council on Educ., 1951.

[13] Thorndike, R. L. *Personnel selection: test and measurement techniques.* New York: Wiley, 1949.

[14] Thorndike, R. L. Review of Davis, F. B. *Item-analysis data: their computation, interpretation, and use in test construction. Psychometrika,* 1947, **12**, 60–61.

Manuscript received 5/16/58

Revised manuscript received 7/14/58

BOOK REVIEW

Oléron, Pierre. *Les Composantes de l'Intelligence d'après les Recherches Factorielles.* Paris: Presse Universitaires de France, 1957. Pp. viii + 517.

One of the important consequences for psychology of the founding of a quantitative science of biological variation by Francis Galton and Karl Pearson was that Charles Spearman, in 1904, could present to the field a paper entitled "General intelligence objectively determined and measured." This paper not only resulted directly in the establishment of factor analysis as a new discipline in psychology, but it is indirectly responsible for *Psychometrika* and, we must acknowledge, for the livelihood of quite a number of people who read this journal. In the more than fifty years of research and controversy which the paper initiated, factor analysis has become widely accepted as a general method for the study of multivariate relationships and has contributed to the theoretical development of modern statistics. At present the original issues in factor analysis are largely resolved and the field now seems to be in a period of consolidation before entering a second phase. M. Oléron has chosen this very opportune time to present a complete review of contributions in the first phase and to suggest possible directions for research in the second. Some indication of the thoroughness of the review is given by the length as well as by the 471 item bibliography and useful descriptive list of 157 psychological tests. The approach is largely historical, somewhat like Dael Wolfle's *Factor Analysis to 1940* (*Psychometric Monograph No. 3*, 1940), and deals with issues and results without going deeply into methodology. The only recent comparable work in English, Sten Henrysson's *Applicability of factor analysis in the behavioral sciences,* (Stockholm, 1957), is less detailed and more technical. A similar but less current work is P. Vernon's, *Structure of Human Abilities.* (New York and London, 1950).

The review is organized under three main headings: the two-factor theory, the multiple factor approach, and results. The first two sections are developed largely in terms of the running controversy between the advocates of the two-factor theory of human ability, with Spearman at the forefront, and the opposition headed by Godfrey Thomson and L. L. Thurstone. The author follows the published record of the dispute closely, reproducing the arguments and counter arguments of the period, but not adding much in the way of original criticism. Especially in certain technical matters, such as the assumptions about the null distribution of tetrad differences and residual correlations used in the tests of the number of significant factors, some definitive criticism based on modern statistical theory would have been valuable. The treatment of the two points of view in the controversy is eminently fair, but it becomes apparent that the author favors the Thurstonian methods because of their greater generality.

The section on results collects the widely scattered literature of factorial studies into a well-organized summary of major findings through 1955. Important factors are described and numerical results quoted in some cases, but extensive numerical tabulation of factor loadings is avoided. Critical comments in this section are more extensive and help the reader considerably in judging to what extent these findings contribute to a coherent description of human abilities.

A main theme throughout the review is the question of what the factorial studies contribute to the concept of intelligence. Factorists generally have tried to avoid discussing intelligence, but the wide influence of the Binet scale and the "intelligence quotient" in mental testing practice makes this difficult. Both Spearman and Thurstone eventually took positions on the question, and they are compared in the review. The author points out that Spearman was forced to admit the existence of group factors, although only as minor sources of test variability. Similarly, the strict multi-factor approach of Thurstone was compromised when the primary abilities were found to be correlated. Thurstone interpreted the second-order factor implied by these correlations as general ability acting in varying degrees through the modality of group factors. He put principal emphasis on the group factors, however, and believed that intelligence would be understood in a useful way when the catalogue of group factors was complete.

But it is the opinion of Oléron that the program of factorial exploration envisioned by Thurstone has become bogged down by the discovery of a seemingly endless number of group factors. Evidence is cited indicating that the appearance of group factors, and more important the variance attributable to them, depends upon the arbitrary choice of test battery composition and cannot be uniquely determined. As a consequence, factors become increasingly fragmented as batteries are more minutely specialized, and there is no obvious way to bring them into a common perspective. Various attempts to organize the group factors according to their importance in practical terms, such as academic success, are reviewed by the author and rejected as inconclusive and not of sufficient theoretical interest. His alternative suggestion is a more experimental approach intended to characterize intelligence not as a catalogue of abilities but as a certain quality of functional relationship among behaviors required in a variety of tasks. Instead of a search for pure tasks representing isolated abilities, as attempted with factor analysis, the study of intelligence would be closer to that of problem solving behavior and might profit from studies of both pathological and normal subjects. This point of view apparently reflects the author's interest in the intellectual capacities of deaf-mutes. His general view is that the factor analytic approach is valuable in identifying the major dimensions of human ability but cannot supply the more detailed factual information necessary to support a substantial theory of intelligence. These facts, he believes, will have to come largely from experimental studies.

University of North Carolina R. Darrell Bock

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

ITEM SELECTION PROCEDURES FOR ITEM VARIABLES WITH A KNOWN FACTOR STRUCTURE*

G. ELFVING
UNIVERSITY OF HELSINKI

R. SITGREAVES AND H. SOLOMON
COLUMBIA UNIVERSITY

This paper discusses the item selection problem when the item responses follow a linear multiple factor model. Because of this restrictive assumption, not too unrealistic in situations such as mental testing, it is possible to select optimal sets of items without going through all possible combinations. A method proposed by Elfving to accomplish this is analyzed and then demonstrated through the use of two illustrations. The common and often used procedure of observing the magnitude of the correlation coefficient as an index in item selection is shown to have some merit in the single-factor case.

The question of type and number of items to be used in a test of mental ability, an attitude scale, a personality inventory, or biographical inventory is a familiar topic in psychological testing. This paper will attempt some resolution of the problem for the restricted situation where the item variables obey a known factor structure. For simplicity the generic term *test* will be taken to represent a collection of items whether dealing with attitude, personality, etc. The classical situation where test response is a linear function of the item responses will be considered. This, of course, may be somewhat unrealistic in some situations, e.g., biographical inventories may call for nonclassical approaches in item selection. On the other hand, even for what may be termed the classical situation the question has not been resolved.

Assume that the test response is to be used as a predictor of a criterion variable. The number and type of items selected for a test, therefore, are governed by the usefulness of the resulting test response for prediction purposes. In the usual techniques of item selection, the size of a correlation coefficient between an item response and the criterion indicates how well the item aids in the prediction made by the test. While the correlation coefficient appears to have pragmatic value for tests of mental ability, it falls short in a number of other testing situations. It should be remembered that the correlation coefficient is a methodological tool borrowed from those who conceptualized its use in anthropometric settings; it may not be efficient in all psychometric situations. A conceptualization of the problem which

---

*Work performed under contract AF 41(657)-244 with the School of Aviation Medicine, Randolph AFB, Texas.

leads to a more general approach to item selection will be considered here.

Suppose for prediction of a criterion z, a large number N of nonrepeatable observations item variables x_i is potentially available. The quantities $x_1 , x_2 , \cdots , x_N ; z$ are assumed to be random variables having joint distributions which are known except perhaps for certain parameters, from past experience or by assumption. In other words, sampling questions are completely disregarded, and we restrict ourselves to a design problem only.

For practical reasons, one wants to base the prediction of z on a restricted number, say $n < N$, of observations x_i . The problem is how to choose them. In a psychological application one can think of the x_i as scores obtained as responses to items on a reading test, and the criterion z as the school grade in which a child will be classified on the basis of a moderate number of items. For concreteness, one can think of N as being of the order 100 to 1000, and n from 10 to 50.

The proposed item selection approach is based on the assumption that the variables $x_1 , x_2 , \cdots , x_N ; z$ have a known factor structure, with a comparatively small number k of common factors; assume that, practically speaking, k ranges from 1 to 5. The assumption of this latent factor universe, which is a restrictive assumption, and how this added information provides for item selection is a departure from other item selection methods. As will be seen, it also provides a rationale in one situation for the correlation coefficient, used so extensively at present. This paper is essentially an analytical and expository account of a method proposed by Elfving [1, 2, 5, 6] and is based on some prior work in different contexts [3, 4].

In the present paper, an explanation of the principles leading to the steps of the method is given, and the method is applied to two sets of data. One set comes from the Educational Testing Service in connection with aptitude testing for law school, where the best two out of six items are to be selected. The other set was artificially constructed; from ten items the best four, the best five, and the best six are, respectively, to be selected. It can easily be demonstrated [1] that the best set of $(n + 1)$ items need not contain the best set of n items. One reason for the scarcity of data on which to employ our item selection procedure is that not only must the factor structure for N items be known completely but the factor loadings for the criterion variable z must be known or guessed realistically. The procedure to be described is also useful for selecting tests from a battery of tests and in these situations, complete factor structures are available. However the language and demonstrations will stress the selection of items for a test.

The Factor Structure and Prediction Criterion

Consider the following factor structure for the x_i and z. Naturally if the items are to have any validity in predicting the criterion, they should be composed of the same common factors and differ at most in factor loadings and specific factors. Thus,

$$x_i = a_{i1}y_1 + \cdots + a_{ik}y_k + \epsilon_i \qquad (i = 1, \cdots, N), \tag{1}$$

$$z = c_1y_1 + \cdots + c_ky_k + \eta, \tag{2}$$

where

(*i*) the *loadings* a_{ij} , c_j are known constants;

(*ii*) the (unobservable) *specific factors* $\epsilon_1, \cdots, \epsilon_N, \eta$ are random variables with mean zero, distributed independently of the unobservable (latent) common factors $y_1, \cdots, y_k$;

(*iii*) the ϵ_i have known covariance matrix (in most of what follows, this will be assumed to be diagonal);

(*iv*) η has variance σ_c^2 and is uncorrelated with the ϵ_i .

For the common factors y_j , consider two different models which, however, lead essentially to the same selection technique;

(*v* a) fixed constants model where the y_i are unknown constants;

(*v* b) random factor model where the y_i are random variables, with mean zero and a known nonsingular covariance matrix T.

In model (*v* a) the y_j may be thought of as factor values pertaining to the particular individual for which z has to be predicted. In (*v* b) the individual is thought of as belonging to a population with known characteristics.

In predicting z, only linear unbiased minimum-variance predictors will be considered; i.e., for any selected set of n items out of N items, a predictor, is taken to be the linear combination

$$\hat{z} = \sum_{i=1}^{n} q_ix_i , \tag{3}$$

which satisfies both $E(\hat{z} - z) = 0$, and $E(\hat{z} - z)^2 =$ minimum. This is a reasonable criterion for estimation quite often used in classical multivariate analysis. Suppose for the selected set of n items in matrix notation,

$$x = Ay + \epsilon, \tag{4}$$

$$z = c'y + \eta, \tag{5}$$

and

$$\Sigma = \text{cov}\ \epsilon = E(\epsilon\epsilon'). \tag{6}$$

Then, in the fixed constants model, the best predictor of z on the basis of x is

$$\hat{z} = c'(A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}x, \tag{7}$$

with prediction variance

$$E(\hat{z} - z)^2 = c'(A'\Sigma^{-1}A)^{-1}c + \sigma_c^2 . \tag{8}$$

In the random factor model, the best predictor of z, on the basis of x, is

$$\hat{z} = c'(T^{-1} + A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}x, \tag{9}$$

with prediction variance

$$E(\hat{z} - z)^2 = c'(T^{-1} + A'\Sigma^{-1}A)^{-1}c + \sigma_c^2 . \tag{10}$$

These results are merely stated since they are well known in multivariate analysis.

The Selection Region

Now consider the selection problem. The problem is to choose n out of N items. There are many possibilities but that set of n items is desired which will minimize the prediction variance $E(\hat{z} - z)^2$. Naturally a procedure is preferred which makes it unnecessary to look at all possible sets of n items to arrive at the proper choice. Considerations of all possibilities can be done for small n and small N, but even when $N = 10$ and $n = 2$, it is quite tedious.

The prediction variance is given, for the two models considered, by (8) and (10). Since σ_c^2 does not depend on the selection, the problem reduces to the minimization of

$$V = c'M^{-1}c, \tag{11}$$

where

$$M = M_0 + A'\Sigma^{-1}A, \tag{12}$$

with $M_0 = 0$ in the parametric model, and $M_0 = T^{-1}$ in the random factor model. The variable elements are, of course, in the $n \times k$ matrix A, and the $n \times n$ matrix Σ, both of which depend on the choice of items.

From now on, assume that the specific factors are uncorrelated, with variances σ_i^2. That is, for a selected set of n items,

$$A'\Sigma^{-1}A = \begin{bmatrix} \sum_{i=1}^{n} \frac{a_{i1}^2}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i1}a_{i2}}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{i1}a_{ik}}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{a_{i1}a_{i2}}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i2}^2}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{i2}a_{ik}}{\sigma_i^2} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \frac{a_{i1}a_{ik}}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i2}a_{ik}}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{ik}^2}{\sigma_i^2} \end{bmatrix}. \tag{13}$$

Denote by a_i row i of A written as a column vector. Then,

$$A'\Sigma^{-1}A = \sum_{i=1}^{n} \frac{a_i a_i'}{\sigma_i^2} = \sum_{i=1}^{n} u_i u_i' , \quad \text{where} \quad u_i = \frac{a_i}{\sigma_i} ; \tag{14}$$

that is,

$$u_iu_i' = \begin{bmatrix} \frac{a_{i1}^2}{\sigma_i^2} & \frac{a_{i1}a_{i2}}{\sigma_i^2} & \cdots & \frac{a_{i1}a_{ik}}{\sigma_i^2} \\ \frac{a_{i1}a_{i2}}{\sigma_i^2} & \frac{a_{i2}^2}{\sigma_i^2} & \cdots & \frac{a_{i2}a_{ik}}{\sigma_i^2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{a_{i1}a_{ik}}{\sigma_i^2} & \frac{a_{i2}a_{ik}}{\sigma_i^2} & \cdots & \frac{a_{ik}^2}{\sigma_i^2} \end{bmatrix} \tag{15}$$

is, for every i, a $k \times k$ matrix of rank 1. The sum $\sum_{i=1}^{n} u_iu_i'$ has for elements the moments of the k-dimensional item population consisting of the points $u_1, \cdots u_n$, each with weight 1.

Returning to the original numbering of the items, $i = 1, \cdots, N$, and denoting by ω the selected set of n subscripts, (12) may now be written

$$M = M_0 + \sum_{\omega} u_iu_i' . \tag{16}$$

The $k \times k$ matrix M is the *information matrix* of the experiment, i.e., of the selected set of observations. The sum $\Sigma u_iu_i'$ represents the information offered by the items in ω. In the random factor model, $M_0 = T^{-1}$ may be said to represent the a priori information contained in the assumption that the factors are random variables with zero means and covariance matrix T. It may be noted, incidentally, that the constant term M_0 in (16) can also be used to take care of any *fixed* source of information, such as possible "compulsory" items.

It is seen that the effect of a particular item-number i, say, depends solely on the vector u_i, i.e., on the reduced loading vector obtained by standardizing the item variable to have variance equal to one. In the fixed constants case, one may multiply the u_i by a common factor without affecting the minimization problem at hand. As a consequence, the item variables may actually be standardized to any common variance, not necessarily unity.

Also note that two items, one with vector u_i and the other with vector $-u_i$, yield the same contribution to the information matrix. For this reason and reasons of symmetry which will become clearer later on, each item will usually be described by means of the *pair* of opposite points $\pm u_i$. These points, in k-space, will be referred to as *item points*.

It is natural to think of the selected set ω of item points as occupying a certain "selection region" S which will, of course, depend in some way on the totality of available item points. Before attacking the question of how to find this region, it may be useful to discuss briefly the cases $k = 1$ and $k = 2$.

When $k = 1$, all symbols in (11) and (16) denote scalars, and the problem reduces to minimizing

$$V = \frac{c^2}{m_0 + \sum_{\omega} u_i^2} \tag{17}$$

by a proper choice of ω. Thus one has simply to select the n items with largest $| u_i |$. If the x_i and u_i, as often in factor analysis, are standardized to variance one, $a_i^2 + \sigma_i^2 = 1$, and hence $u_i^2 = a_i^2/\sigma_i^2 = a_i^2/(1 - a_i^2)$. The items with largest $| u_i |$ are then the same as those with largest $| a_i |$, i.e., those having largest loadings with respect to the single common factor. In this case, however,

$$\frac{a_i^2}{1 - a_i^2} = \frac{\rho_{x_i y_1}^2}{1 - \rho_{x_i y_1}^2} , \tag{18}$$

so that this procedure is equivalent to picking the x_i having the highest absolute correlation with y_1. This procedure is usually followed by psychologists, except that the latent factor is not available. Therefore, resort is made to a manifest equivalent, usually observed total test score.

In the case of $k = 2$, it is clear, in general, that if an item point happens to lie precisely on the straight line determined by the vector c, say $u_1 = \lambda_1 c$, then

$$\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 + \epsilon_1 \\ &= \sigma_1\lambda_1(c_1y_1 + c_2y_2) + \epsilon_1 , \end{aligned} \tag{19}$$

and $x_1/\sigma_1\lambda_1$ provides an unbiased estimate of $c'y$. The variance of this estimate is $\sigma_1^2/(\sigma_1\lambda_1)^2 = 1/\lambda_1^2$. Accordingly, item points along the line c may be expected to contribute more to the estimation of $c'y$, and hence to the prediction of z, the farther out they are along the line.

Similarly, suppose there are two item points located more or less symmetrically with respect to the line c. For example, suppose

$$\begin{aligned} u_1 &= \left(\frac{\lambda_{c_1}}{\sqrt{c_1^2 + c_2^2}} + kc_2 \qquad \frac{\lambda_{c_2}}{\sqrt{c_1^2 + c_2^2}} - kc_1\right), \\ u_2 &= \left(\frac{\lambda_{c_1}}{\sqrt{c_1^2 + c_2^2}} + kc_2 \qquad \frac{\lambda_{c_2}}{\sqrt{c_1^2 + c_2^2}} + kc_1\right). \end{aligned} \tag{20}$$

Then

$$\frac{1}{2}\left(\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2}\right) = \frac{\lambda}{\sqrt{c_1^2 + c_2^2}}\,(c_1y_1 + c_2y_2) + \frac{1}{2}\left(\frac{\epsilon_1}{\sigma_1} + \frac{\epsilon_2}{\sigma_1}\right), \tag{21}$$

and

$$\frac{\sqrt{c_1^2 + c_2^2}}{2\lambda}\left(\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2}\right) \tag{22}$$

is an unbiased estimate of $c'y$ with variance

$$\frac{(c_1^2 + c_2^2)}{2\lambda^2}. \tag{23}$$

Since λ is the distance from the origin to the point of intersection of the c-vector and the line joining u_1 and u_2, it is clear again that the two item points, after the elimination of the orthogonal component, provide an estimator of $c'y$ which is better the farther off in the $\pm$ c-direction they lie.

These heuristic remarks make it plausible that the selection region S will have to comprise the outer parts of k-space with regard to the directions $\pm$ c and will have to fulfill the additional requirement that the item points in S should in some way balance each other with respect to that direction. It turns out [1, 2, 5, 6] that S may be taken to consist of two symmetrical half-spaces (a twin half-space) bounded by two parallel planes $t'u = \pm h$; the direction t of their common normal will depend in a certain manner on the item points included in the selection region, the latter becoming in this way implicitly determined.

In order to formulate the result just sketched, it is necessary to introduce a continuation device which will lead to a simple optimizing criterion. This will greatly facilitate the solution of our problem without essentially affecting the practical application. For this purpose consider, instead of the previous M defined by (16), the generalized information matrix

$$M = M_0 + \sum_{i=1}^{N} p_i u_i u_i', \tag{24}$$

where the allocation vector $p = (p_1 \cdots p_N)$ is subject to the restrictions

$$0 \leq p_i \leq 1, \qquad \sum_1^N p_i = n. \tag{25}$$

Obviously, the set of matrices given by (24) and (25) contains the set of matrices (16). For any particular p, those items for which $p_i = 1$ will be referred to as *totally selected*, those for which $p_i = 0$, as *nonselected*, and those for which $0 < p_i < 1$, as *fractionally selected*.

The following interpretation of the fractional p_i (in the parametric case) may be instructive. Imagine for a moment that the observations x_i may be independently repeated, each of them at most r times, and assume that a total of rn observations is allowed. Let $n_1, \cdots, n_N$ ($0 \leq n_i \leq r$; $\sum n_i = rn$) be the number of times that the different observations are repeated. The information matrix of the resulting experiment may be written

$$M = \Sigma n_i u_i u_i' = r \Sigma p_i u_i u_i', \tag{26}$$

where the $p_i = n_i/r$ vary from 0 to 1 through multiples of $1/r$, subject to the condition $\Sigma p_i = n$. Since the factor r in M is obviously irrelevant to the minimization of (11), for large r, the problem is essentially a modified selection version of that formulated above. In particular, if $n = 1$, and r is large, one is concerned with the allocation problem treated by Elfving in [3]; the earlier problem thus appears a special case of the present one.

After these preparations, one may state the following propositions [2, 5].

THEOREM 1. *The scalar* V, *defined by*

$$V = c'M^{-1}c, \qquad \textit{where} \quad M = M_0 + \sum_{i=1}^{N} p_i u_i u_i' \tag{27}$$

as a function of the vector p, *has a minimum on the domain (25). In order for the allocation vector* p *to yield this minimum, it is necessary and sufficient that, for a certain number* $h > 0$,

$$p_i = \begin{cases} 1 & \textit{whenever} \quad | c'M^{-1}u_i | > h \\ 0 & \textit{whenever} \quad | c'M^{-1}u_i | < h. \end{cases} \tag{28}$$

Moreover, there exists always a minimizing p *with at most* k *fractional components.*

Recall that M is a $k \times k$ matrix, hence $c'M^{-1}$ is a k-dimensional row vector, and $c'M^{-1}u_i$ is a linear form in the components of the vector u_i. The content of the theorem is that the selection region consists mainly of that part of k-space which lies outside two parallel hyperplanes $c'M^{-1}u = \pm h$; the item points lying in the boundary planes will have to be totally selected, fractionally selected, or nonselected, as the case may be. Since the fractional p_i (when they are taken to be as few as possible) can total at most k, there will be from $n - k + 1$ to n totally selected items. In practice, one may round off some or all of the fractional p_i to unity, i.e., select the corresponding items on an equal basis with the rest. In the latter case, at the expense of making at most $k - 1$ more observations than originally planned, one will be sure to achieve a variance V not exceeding the smallest one that could be attained by any total selection of n items. In those cases where the optimal p will contain only 1's and 0's, an exact solution of the original discrete problem is provided.

For a proof of Theorem 1 in the parametric case (the random factor case goes quite similarly) the reader is referred to [2, 5]. It should be noted that the theorem gives a necessary and sufficient criterion for optimum solutions, but no method for finding such a solution. When $k = 2$ and n is small, a graphic picture of the item points may lead to a good guess. For more complex situations, a method has been suggested [1, 6] based on the idea that the population of item points may be approximately described by a k-dimensional normal distribution with the same second-order moments. The method may be condensed into the following practical rule.

(*i*) Find the matrix Λ with elements

$$\lambda_{jh} = \sum_{i=1}^{N} u_{ij} u_{ih} \qquad (j, h = 1, \cdots, k). \tag{29}$$

(*ii*) Find the vector $\gamma = \Lambda^{-1}c$, i.e., solve the equations

$$\begin{aligned} \lambda_{11}\gamma_1 + \cdots + \lambda_{1k}\gamma_k &= c_1 , \\ &\cdots \\ \lambda_{k1}\gamma_1 + \cdots + \lambda_{kk}\gamma_k &= c_k . \end{aligned} \tag{30}$$

(*iii*) Find, for each i, the quantity

$$w_i = \gamma' u_i = \sum_j \gamma_j u_{ij} \tag{31}$$

and select the n items with largest $| w_i |$.

The selection found in this way should provide a good first guess and, of course, may be checked by means of Theorem 1. If the criterion is not fulfilled, one may try to improve the solution by exchanging one or more of the selected item points for others in the neighborhood of the boundary planes. An artificial example of such a procedure is given in a later section.

A Realistic Example when $k = 2$

To illustrate the meaning and use of Theorem 1, two examples will be considered, both of which involve a two-factor structure and assume a fixed constants model. The first example is based on data made available by the Educational Testing Service. These data resulted from responses to six items used to measure aptitude for success in law school and a criterion variable which measured success in law school. A factor analysis of the six items and the criterion variable resulted in two common factors plus specific factors. For this illustration,

$$\begin{aligned} x_i &= a_i' y + \epsilon_i \qquad (i = 1, 2, \cdots, 6), \\ z &= c' y + \eta, \end{aligned} \tag{32}$$

where

$$\begin{aligned} a_1' &= (.848 \quad .225), \quad a_2' = (.833 \quad .195), \quad a_3' = (.840 \quad .122), \\ a_4' &= (.481 \quad -.216), \quad a_5' = (.647 \quad -.172), \quad a_6' = (.869 \quad .204), \\ c' &= (.641 \quad -.205). \end{aligned} \tag{33}$$

The (unobservable) specific factors $\epsilon_1 , \cdots , \epsilon_6 ; \eta$ are assumed to be independently distributed random variables with mean zero. The variances of $\epsilon_1 , \cdots , \epsilon_6$ are

$$\begin{aligned} \sigma_1^2 &= .230, \quad \sigma_2^2 = .268, \quad \sigma_3^2 = .280, \\ \sigma_4^2 &= .722, \quad \sigma_5^2 = .552, \quad \sigma_6^2 = .203. \end{aligned} \tag{34}$$

The vector of (unobservable) common factors $y' = (y_1 y_2)$ is assumed to be a vector of unknown constants.

Then, if $u_i = a_i/\sigma_i$,

$$u_1' = (1.767 \quad .469),\quad u_2' = (1.608 \quad .376),\quad u_3' = (1.588 \quad .231),$$
$$u_4' = (.566 \quad -.254),\ u_5' = (.871 \quad -.231),\ u_6' = (1.927 \quad .452); \tag{35}$$

and

$$u_1u_1' = \begin{bmatrix} 3.122 & .829 \\ .829 & .220 \end{bmatrix},\quad u_2u_2' = \begin{bmatrix} 2.586 & .605 \\ .605 & .141 \end{bmatrix},$$
$$u_3u_3' = \begin{bmatrix} 2.522 & .367 \\ .367 & .053 \end{bmatrix},\quad u_4u_4' = \begin{bmatrix} .320 & -.144 \\ -.144 & .065 \end{bmatrix}, \tag{36}$$
$$u_5u_5' = \begin{bmatrix} .759 & -.201 \\ -.201 & .053 \end{bmatrix},\quad u_6u_6' = \begin{bmatrix} 3.713 & .871 \\ .871 & .204 \end{bmatrix}.$$

The item points $u_1, \cdots, u_6$, and the vector c are shown graphically in Figure 1.

For a given set of x's, say $x_{i_1}, \cdots, x_{i_n}$, the best predictor of z on the basis of x is

$$z = c'M_\omega^{-1}U_\omega x^*, \tag{37}$$

where ω denotes the selected set of subscripts, and

$$M_\omega = \sum_\omega u_iu_i'; \qquad U_\omega = (u_{i_1} \quad u_{i_2} \quad \cdots \quad u_{i_n}),\text{ a } 2 \times n \text{ matrix};$$
$$x^* = \left(\frac{x_{i_1}}{\sigma_{i_1}} \quad \frac{x_{i_2}}{\sigma_{i_2}} \quad \cdots \quad \frac{x_{i_n}}{\sigma_{i_n}}\right).$$

The variance of the estimate is

$$E(z - \hat{z})^2 = c'M^{-1}c + \sigma_c^2. \tag{38}$$

M_ω may be considered the information matrix for the selected set of observations.

Consider the generalized information matrix

$$M_p = \sum_{i=1}^{6} p_iu_iu_i', \tag{39}$$

where the allocation vector $p' = (p_1 \quad p_2 \quad \cdots \quad p_6)$ is subject to the restrictions

$$0 \le p_i \le 1, \qquad \sum_{i=1}^{6} p_i = 2. \tag{40}$$

That is, two out of the six items are desired. Then, Theorem 1 states that

$$V = c'M_p^{-1}c, \tag{41}$$

as a function of the vector p, has a minimum on the domain (40). In order for the allocation vector $p = p^*$ to yield this minimum, it is necessary and sufficient that for a certain number $h > 0$

$$p_i = \begin{cases} 1 & \text{whenever} \quad |c' M_{p^*}^{-1} u_i| > h \\ 0 & \text{whenever} \quad |c' M_{p^*}^{-1} u_i| < h, \end{cases}$$

where p^* is the minimizing vector. That is, the wholly selected item points lie outside of two parallel lines which are symmetric with respect to the origin, while the totally nonselected item points lie between these lines. (Totally selected and totally nonselected item points may also lie on the boundaries, i.e., parallel lines.) Any fractionally selected item, say n_i, must be on one of the lines defined by $| c' M_{p^*}^{-1} u_i | = h$. At most, two items may need to be fractionally selected.

Suppose in the present example, where $n = 2$, we first limit ourselves to picking the best two wholly selected items; that is, the possibility of fractionally selected items is not permitted. If one examines the 15 possible combinations of the two items in pairs, it is seen that items 4 and 5 are best, i.e., they yield the estimate with the smallest variance. Writing $p_0 = (0, 0, 0, 1, 1, 0)$, the variance of this estimate is

$$c' M_{p_0}^{-1} c = .381. \tag{42}$$

Theorem 1 states that if, in fact, p_0 is the minimizing vector when fractionally selected items are admitted, there is a number h such that

$$| c' M_{p_0}^{-1} u_4 | > h, \quad | c' M_{p_0}^{-1} u_5 | > h, \quad | c' M_{p_0}^{-1} u_j | < h \quad (j = 1, 2, 3, 6)\cdot \tag{43}$$

However,

$$\begin{aligned} | c' M_{p_0}^{-1} u_1 | &= 1.043, \quad | c' M_{p_0}^{-1} u_2 | = .950, \quad | c' M_{p_0}^{-1} u_3 | = .939, \\ | c' M_{p_0}^{-1} u_4 | &= .337, \quad | c' M_{p_0}^{-1} u_5 | = .517, \quad | c' M_{p_0}^{-1} u_6 | = 1.141. \end{aligned} \tag{44}$$

It follows that p_0 is not the minimizing vector, and it will be necessary to consider fractionally selected items.

Since items 4 and 5 are the best pair of items (the variance of the estimate based on items 2 and 6, for example, is 447.01), it seems reasonable that these items be included when fractional selection is allowed. The theorem indicates that there must be a pair of parallel lines such that the two fractionally selected items lie on the line.

It is seen in Figure 1 that if the two parallel lines are drawn through items 3 and 4, item 5 lies outside the lines and items 1, 2, and 6 lie between them. Thus one is led to attempt a fractional solution of the form

$$p^* = (0 \quad 0 \quad r \quad 1 - r \quad 1 \quad 0). \tag{45}$$

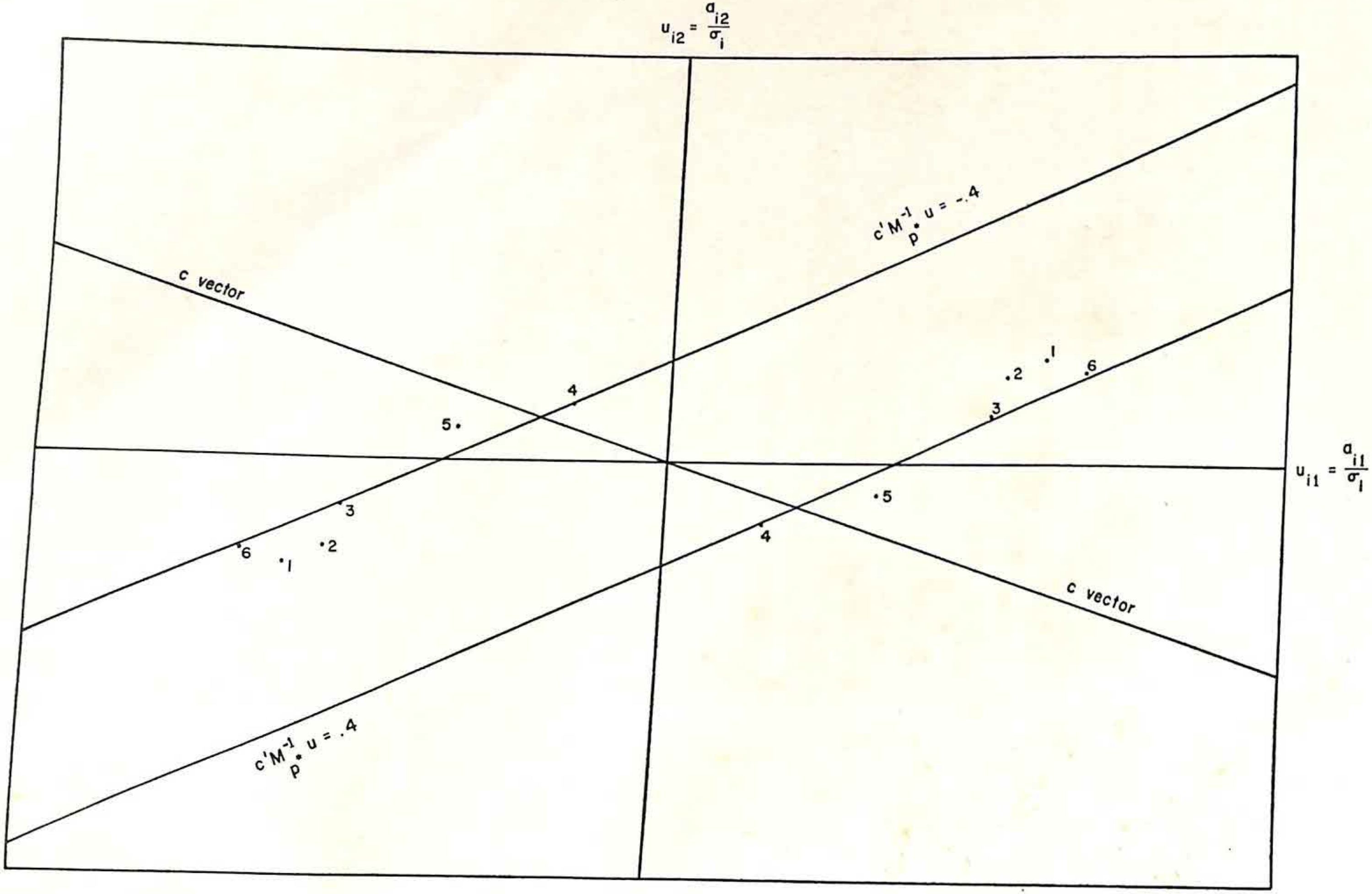

$u_{i2} = \frac{a_{i2}}{\sigma_i}$
$u_{i1} = \frac{a_{i1}}{\sigma_i}$
c vector
c vector
$c'M_{p\cdot}^{-1} u = -.4$
$c'M_{p\cdot}^{-1} u = .4$
1
2
3
4
5
6
1
2
3
4
5
6

FIGURE 1
Illustration of Solution for Realistic Data: Section 4

Determine r such that

$$c'M_{p^*}^{-1}u_3 = c'M_{p^*}^{-1}u_4 .$$

That is,

$$\begin{aligned} &[.641 \quad -.205]\begin{bmatrix} .118 - .012r & .345 - .511r \\ .345 - .511r & 1.079 + 2.202r \end{bmatrix}\begin{bmatrix} 1.588 \\ .231 \end{bmatrix} \\ &= [.641 \quad -.205]\begin{bmatrix} .118 - .012r & .345 - .511r \\ .345 - .511r & 1.079 + 2.202r \end{bmatrix}\begin{bmatrix} .566 \\ -.254 \end{bmatrix}. \end{aligned} \tag{46}$$

From this equation $r = .019$.

It follows that

$$p^* = (0 \quad 0 \quad .019 \quad .981 \quad 1 \quad 0); \tag{47}$$

$$M_{p^*} = \begin{bmatrix} 1.1208 & -.3353 \\ -.3353 & .1182 \end{bmatrix}; \tag{48}$$

$$c'M_{p^*}^{-1} = (.3506 \quad -.7399); \tag{49}$$

$$\begin{aligned} c'M_{p^*}^{-1}u_1 &= .273, & c'M_{p^*}^{-1}u_2 &= .286, & c'M_{p^*}^{-1}u_3 &= .386, \\ c'M_{p^*}^{-1}u_4 &= .386, & c'M_{p^*}^{-1}u_5 &= .476, & c'M_{p^*}^{-1}u_6 &= .341. \end{aligned} \tag{50}$$

Thus, the conditions of the theorem are satisfied, and p^* is the minimizing vector. The variance of the associated estimate is .376. When this is contrasted with .381, the variance when item 4 and item 5 are used totally, a slight improvement by the use of fractional allocation is seen.

It should be remembered, however, that the modification of the problem by admitting fractional allocation is an ad hoc device, applying primarily when n is considerably larger than k; its justification is to be found in the paragraph following Theorem 1. Thus, in the present example, the *practical* problem is to find the best *integral* solution, i.e., the best pair of items. The fractional solution (47) suggests that items 4 and 5 comprise the best choice, and this is actually the case, as has been checked by comparing the variances corresponding to all possible choices of two items.

An Artificial Example when $k = 2$

For the selection of two out of six items, in the previous example the variance of z actually was computed for all 15 pairs in the two-factor structure; it was found that one could improve on the best pair by fractional allocation. It was shown how Theorem 1, combined with graphical considerations, could be used to select the best pair of items without going through all possible pairs. This becomes especially important when N and n both increase and thus rule out the examination of all possibilities. It also becomes important when k, the number of common factors, increases.

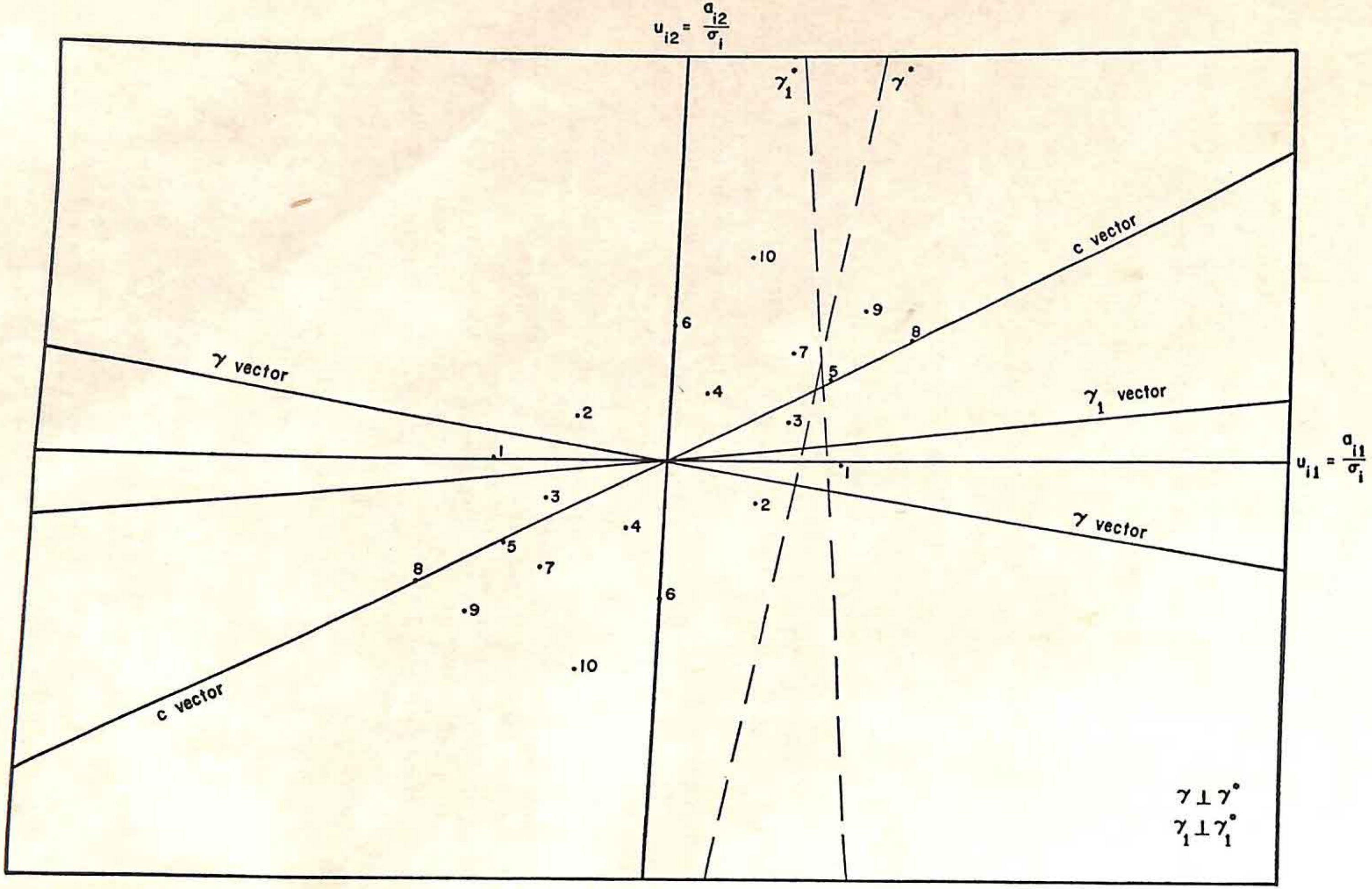


FIGURE 2
Illustration of Solution for Artificial Data: Section 5

The application of the theorem in a more complex situation will be demonstrated by the following artificial example. The problem is to obtain from the 10 items, first the optimal sets of 4 items, then of 5 items, and finally of 6 items. A common factor space of two dimensions is assumed. Thus,

$$\begin{aligned} x_i &= a_i'y + \epsilon_i \qquad (i = 1, \cdots, 10), \\ z &= c'y + \eta, \end{aligned} \tag{51}$$

where $c' = (2 \quad 1)$. Assume

$$\begin{aligned}
u_1' &= (\ \ 4 \quad 0), & u_6' &= (0 \quad 3), \\
u_2' &= (-2 \quad 1), & u_7' &= (3 \quad 3), \\
u_3' &= (\ \ 3 \quad 1), & u_8' &= (6 \quad 3), \\
u_4' &= (\ \ 1 \quad 2), & u_9' &= (5 \quad 4), \\
u_5' &= (\ \ 4 \quad 2), & u_{10}' &= (2 \quad 5);
\end{aligned} \tag{52}$$

$$\begin{aligned}
u_1u_1' &= \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix}, & u_6u_6' &= \begin{bmatrix} 0 & 0 \\ 0 & 9 \end{bmatrix}, \\
u_2u_2' &= \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}, & u_7u_7' &= \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}, \\
u_3u_3' &= \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}, & u_8u_8' &= \begin{bmatrix} 36 & 18 \\ 18 & 9 \end{bmatrix}, \\
u_4u_4' &= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, & u_9u_9' &= \begin{bmatrix} 25 & 20 \\ 20 & 16 \end{bmatrix}, \\
u_5u_5' &= \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix}, & u_{10}u_{10}' &= \begin{bmatrix} 4 & 10 \\ 10 & 25 \end{bmatrix}.
\end{aligned} \tag{53}$$

Applying the approximate method for complex situations, presented earlier,

$$\Lambda = \sum_{i=1}^{10} u_iu_i' = \begin{bmatrix} 120 & 68 \\ 67 & 88 \end{bmatrix}. \tag{54}$$

Then

$$\Lambda^{-1} \sim \begin{bmatrix} 78 & -68 \\ -68 & 120 \end{bmatrix} \tag{55}$$

and

$$\gamma' = c'\Lambda^{-1} \sim (88 \qquad -16). \tag{56}$$

In (55) and (56) we need not be concerned with proportionality factors.

The item points are plotted graphically in Figure 2. Plotting the direction of γ in Figure 2, and imagining a straight line perpendicular to it moving from right to left, then the order in which the item points are swept over by this line indicates their order of preference in the first approximation.

If $n = 4$, i.e., 4 items are to be selected, one is led to the selection I_1 : 1, 5, 8, 9. Forming the corresponding product sums,

$$M_1^{-1} \sim \begin{bmatrix} 29 & -46 \\ -46 & 93 \end{bmatrix}, \qquad \gamma_1' \sim (12 \quad 1). \tag{57}$$

In the direction of γ_1 , the four outermost item points are still numbers 1, 5, 8, 9. The theorem shows, then, that choice I is optimal.

Next, take $n = 5$. Using the approximate selection determined by the γ-direction, one tries the set II_1 : 1, 3, 5, 8, 9. From the corresponding product sums, $\gamma_1' \sim (11 \quad 4)$. In this direction, however, the five outermost item points are II_2 : 1, 5, 7, 8, 9. On the other hand, the latter set yields the direction $\gamma_2' \sim (21 \quad -8)$, and the five outermost points 1, 3, 5, 8, 9, i.e., the set first attempted. This mutuality makes it plausible that a selection of form II_3 : 1, 3*, 5, 7*, 8, 9 (* indicating fractional selection) might give an optimal solution of the problem. Introducing unknown weights p_3 , p_7 with sum one,

$$M_3 \sim \begin{bmatrix} 93 + 9p_3 + 9p_7 & 46 + 3p_3 + 9p_7 \\ 46 + 3p_3 + 9p_7 & 29 + p_3 + 9p_7 \end{bmatrix}. \tag{58}$$

In order for the points 3, 7 to lie on the boundary line of the selection region, the vector $\gamma = M^{-1}c$ must have direction (1, 0), which gives the condition

$$-2(46 + 3p_3 + 9p_7) + (93 + 9p_3 + 9p_7) = 0; \tag{59}$$

hence (noting that $p_3 + p_7 = 1$), $p_3 = 2/3$, $p_7 = 1/3$. Thus the optimum allocation vector is $p^* = (1, 0, 2/3, 0, 1, 0, 1/3, 1, 1, 0)$. In practice, one must make a choice between selections II_1 and II_2 . The corresponding variances $V = c'M^{-1}c$ are (note that the factors $1/|M|$ cannot be omitted at this point)

$$V_1 = \frac{30(2^2) - 2(49)(2)(1) + 102(1^2)}{102(30) - 49(49)} = \frac{26}{659} = 0.03945,$$

$$V_2 = \frac{38(2^2) - 2(55)(2)(1) + 102(1^2)}{102(38) - 55(55)} = \frac{34}{851} = 0.03995.$$

Thus, there will be a slight preference for the former choice.

Finally, if one wishes to include $n = 6$ items, it seems reasonable, from the figure, to try the set III: 1, 3, 5, 7, 8, 9. It leads to $\gamma_1' \sim (20 \quad -5)$. In this case, the moving boundary line will hit points 2 and 7 at the same time. Nevertheless, the set III still fulfills the condition of the theorem, and hence is optimal. Note that in this example, of the two item points on the boundary line, item 7 has weight 1, and item 2 has weight 0.

Early in the paper it was stated that the optimal $(n + 1)$ set of items need not completely contain the optimal set of n items. However in many practical situations this will occur and the present exercise demonstrates this event.

REFERENCES

[1] Elfving, G. Selection of item variables for prediction. Report No. 56-91. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, August, 1956.

[2] Elfving, G. Further contributions to the theory of item selection. Report No. 57-97. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, June, 1957.

[3] Elfving, G. Optimum allocation in linear regression theory. *Ann. math. Statist.*, 1952, **23**, 225-262.

[4] Elfving, G. Geometric allocation theory. *Skandinavisk Aktuarietidskrift.*, 1954, **37**, 170-190.

[5] Elfving, G. A selection problem in experimental design, *Societas Scientarium Fennica, Commentationes Physico-Mathematicae*, 1957, **20**, Ch. 2, pp. 1-10.

[6] Elfving, G. Selection of nonrepeatable observations for estimation. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: Univ. California Press, 1955. Pp. 69-75.

Manuscript received 8/18/58

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A METRIC AND AN ORDERING ON SETS

FRANK RESTLE

MICHIGAN STATE UNIVERSITY

Beginning with sets of arbitrary elements, concepts of distance and betweenness of sets are defined. Since betweenness as defined is not transitive, an investigation is made of the conditions which ensure desirable regularity. It is found that a straight line or linear array of sets is a generalization of nested sets (Guttman scales). Close relationships among the notions of distance, betweenness, and linear arrays are demonstrated. Parallel and perpendicular arrays, dimensions, and multidimensional spaces are characterized.

Basic to many psychological discussions is the concept of similarity, which is used to arrange objects or events. Two things which are quite similar are psychologically close together, and two things which are quite dissimilar are psychologically distant. This interpretation of dissimilarity as distance is the basis of numerous attempts to give a simplifying, quantitative structure to psychological data in psychophysics, learning, and other fields.

In contrast to these metric, quantitative developments are more qualitative discussions of similarity and the arrangement of objects, using the idea of common elements to account for similarity. In learning, such views are common in current mathematical theories [3]. Some recent efforts have been made to generate a metric analysis from the set-theoretic considerations of learning theory [2, 3] and psychophysics [4]. The present paper shows how a distance and ordering can be defined on sets, and generates the beginnings of the resulting geometry.

The arbitrary elements to be considered have no individual locations or arrangement except with regard to the sets of which they are members. Thus, the foundation is entirely qualitative. The purpose is to determine what internal evidence would justify saying that a sequence of sets is ordered, and to construe a concept of *distance* between sets. The conceptual apparatus used is elementary set theory.

The following mathematical concepts and notations are used.

$\mathcal{U}$	the universe.
$\emptyset$	the empty set.
$\cup$	union or set sum; $A \cup B$ is the set containing everything in either A or B or both.
$\cap$	intersection or set product; $A \cap B$ is the set containing just those elements common to both A and B.

$^-$ set-complement; $\bar{A}$ is the set containing all elements of $\mathcal{U}$ which are not in A.

$\subset$ the relation of inclusion; $A \subset B$ if and only if all members of A are also members of B.

m a measure function; if S^* is a set of sets $\{S_1, S_2, \cdots\}$ and m is a measure function, then the following axioms hold.

$M1$: for all S_i, $m(S_i) \geq 0$.

$M2$: $m(\emptyset) = 0$.

$M3$: if $S_i \cap S_j = \emptyset$, then $m(S_i \cup S_j) = m(S_i) + m(S_j)$.

$\{a, b\}$ is the set whose only members are a and b. In general, entities enclosed in curved brackets are the members of a set.

$\langle a, b \rangle$ the ordered set whose first member is a and second member is b. In general, entities enclosed in angle brackets are members, in the indicated order, of an ordered set. Two ordered sets are identical only if they have the same elements in the same order. (Unordered sets are identical if they have the same members—order is immaterial.)

Consider a universe $\mathcal{U}$ of arbitrary elements, which may be interpreted as the universe of stimulus elements, cues, etc., in learning or perceptual models, or which might also be the set of responses to test items, or the population of subjects, etc., in other applications.

Certain subsets of $\mathcal{U}$ can be distinguished by the kind of observation used in the particular experiment. Therefore consider a set $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$ of subsets of $\mathcal{U}$. A necessary addition is a notion of magnitude of a set, which may be the number of elements in the set, or (if the elements receive different weights) the sum of weights of elements, or which may be a more abstract idea of magnitude. The function m is a measure function on the sets in $\mathcal{S}$, and also on any sets which can be formed from members of $\mathcal{S}$ by the operations of union, intersection, and complementation performed a countable number of times. Let this larger set be $\mathcal{S}^*$. This ensures that m will be defined on all the sets in which we have interest.

Metric of Dissimilarity

Let the dissimilarity or *distance* between two sets S_i and S_j be written D_{ij}. For D to be a metric distance, it must satisfy the following metric axioms.

Axiom 1. $$D_{ij} = 0 \quad \text{if and only if} \quad i = j.$$

Axiom 2. $$D_{ij} \geq 0.$$

Axiom 3. $$D_{ij} = D_{ji}.$$

Axiom 4. $$D_{ij} + D_{jk} \geq D_{ik}.$$

Within the present system, the distance D_{ij} must be defined in terms of the sets S_i and S_j, combinations produced by intersection, union, and complementation, and the measure function m. The nature of the elements in $\mathcal{U}$ does not in general give a concept of distance between elements upon which to build the concept of distance between sets.

Other things equal, the degree of dissimilarity between two sets depends on the measure of noncommon elements, i.e., the symmetric set difference. The first proposal of this study is to let the measure of the symmetric set difference be the distance metric.

DEFINITION 1. $D_{ij} = m\,[(S_i \cup S_j) \cap \overline{(S_i \cap S_j)}]$.

The expression on the right is, of course, the symmetric set difference.

THEOREM 1. *D_{ij} is a metric.*

PROOF. $S_i \cup S_i = S_i \cap S_i = S_i$, whence $D_{ii} = m(S_i \cap \bar{S}_i) = m(\phi) = 0$. If m vanishes only for the empty set, D satisfies Axiom 1 of a metric function. Axiom 2 is satisfied because measure functions have nonnegative values, and Axiom 3 is satisfied because of the symmetry of the expression in Definition 1. Axiom 4, the triangle inequality, is shown to hold by partitioning $\mathcal{U}$ into the following eight cells (see Figure 1):

$$S_i \cap S_j \cap S_k = A_1 \;; \qquad S_i \cap S_j \cap \bar{S}_k = A_2 \;;$$
$$S_i \cap \bar{S}_j \cap S_k = A_3 \;; \qquad \bar{S}_i \cap S_j \cap S_k = A_4 \;;$$
$$S_i \cap \bar{S}_j \cap \bar{S}_k = A_5 \;; \qquad \bar{S}_i \cap S_j \cap \bar{S}_k = A_6 \;;$$
$$\bar{S}_i \cap \bar{S}_j \cap S_k = A_7 \;; \qquad \bar{S}_i \cap \bar{S}_j \cap \bar{S}_k = A_8 \,.$$

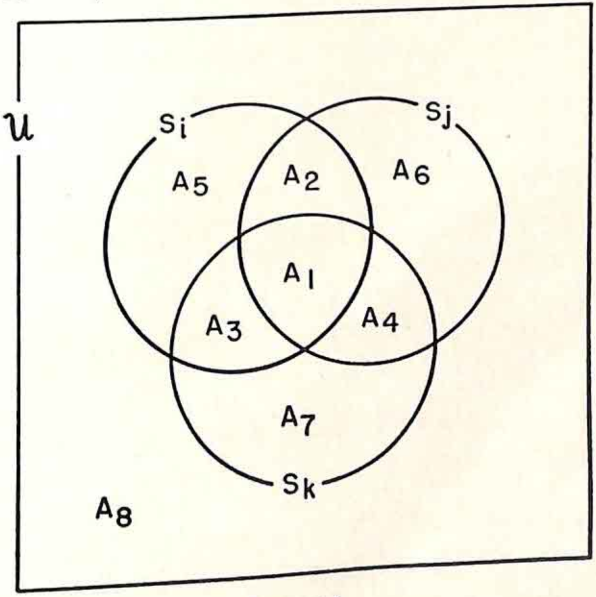


FIGURE 1

Three sets, S_i , S_j , and S_k , in a universe $\mathcal{U}$, with the cells of the resulting partition

For compactness in what follows, let $m(A_i) = a_i$. Now,

$$D_{ij} = a_3 + a_4 + a_5 + a_6 ,$$
$$D_{jk} = a_2 + a_3 + a_6 + a_7 ,$$
$$D_{ik} = a_2 + a_4 + a_5 + a_7 ;$$

whence

$$D_{ij} + D_{jk} = D_{ik} + 2a_3 + 2a_6 ,$$

and

$$D_{ij} + D_{\ k} \geq D_{ik} . \qquad \text{Q.E.D.}$$

An Ordering on Sets

Since the elements of $\mathfrak{U}$ are arbitrary and not already located in a space, or otherwise ordered, an ordering of sets must be defined in terms of the membership of the sets. It appears better to be unrealistically restrictive than vague in the present analysis.

Consider first what it means to say that S_j is *between* S_i and S_k . Two conditions seem sufficient to justify such a description; first that S_i and S_k have no common members which are not also in S_j , and second, that S_j have no unique members which are in neither S_i nor S_k . Figure 2a shows an S_j which is between S_i and S_k in this sense. Figure 2b shows an S_j' which fails to meet the first condition. Figure 2c shows an S_j'' which fails to meet the second condition.

This concept can be written as the following definition.

DEFINITION 2. S_j is between S_i and S_k (written b_{ijk}) if and only if

(i) $S_i \cap \bar{S}_j \cap S_k = \emptyset$ (A_3 in Figure 1 is empty),
(ii) $\bar{S}_i \cap S_j \cap \bar{S}_k = \emptyset$ (A_6 in Figure 1 is empty).

A major justification of definitions 1 (the metric) and 2 (betweenness) is found in the following theorem.

THEOREM 2. *If b_{ijk} then $D_{ij} + D_{jk} = D_{ik}$; if m vanishes only for the empty set then the converse is true.*

PROOF. From the proof of Theorem 1,

$$\begin{aligned} D_{ij} + D_{jk} &= D_{ik} + 2a_3 + 2a_6 \\ &= D_{ik} + 2m(S_i \cap \bar{S}_j \cap S_k) + 2m(\bar{S}_i \cap S_j \cap \bar{S}_k). \end{aligned}$$

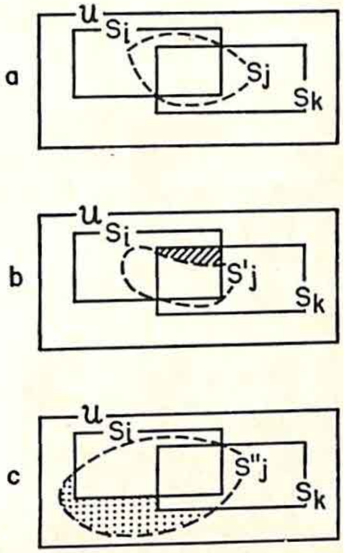


FIGURE 2

An example of ordered sets (a) and two examples of unordered sets (b) and (c)

By hypothesis, b_{ijk}, the sets $S_i \cap \bar{S}_j \cap S_k$ and $\bar{S}_i \cap S_j \cap \bar{S}_k$ are empty, whence $D_{ij} + D_{jk} = D_{ik}$. The converse is obvious if $a_3 = a_6 = 0$ implies that A_3 and A_6 are empty. Q.E.D.

Thus, b_{ijk} means that, in an abstract sense, S_j lies on a straight line between S_i and S_k, in that a path from S_i to S_k through S_j is just as short as any path. A preliminary investigation has shown, however, that one does not easily build a rigid metric space from the relation b. The main trouble is that b is not transitive in what would be a most useful way. The problems are shown in the two counter examples proving Theorem 3.

THEOREM 3. *It is not the case that* b_{ijk} *and* b_{jkm} *implies* b_{ijm}.

PROOF. Two counter examples are given because of their intrinsic interest.

First counter example: let $S_i = \{1, 2, 3\}$, $S_j = \{2, 3, 4\}$, $S_k = \{3, 4, 5\}$, and $S_m = \{4, 5, 1\}$. Then b_{ijk} and b_{jkm}, but not b_{ijm} because $S_i \cap \bar{S}_j \cap S_m = \{1\}$, whereas it must be empty for b_{ijm} to hold.

Second counter example: let $S_i = \{1, 2\}$, $S_j = \{2, 3\}$, $S_k = \{3, 4\}$ and $S_m = \{4, 5\}$. Then b_{ijk} and b_{jkm}, but not b_{ijm} because $\bar{S}_i \cap S_j \cap \bar{S}_m = \{3\}$, whereas it must be empty for b_{ijm} to hold. Q.E.D.

These two examples correspond to what may be considered psychological reality. The first counter example shows a case of *looping* in which a series of neighboring sets, each between its two neighbors, may lead around and back to the starting set. This is the apparent structure of the color circle, the array of hues from red to yellow to green to blue to purple and back to red. In the second counter example, one detours from S_i to S_m through some sets which may be quite unlike either. For example, a series of drawings might be made, varying from starkly simple to profusely ornate. Drawings in the middle of the series could have psychological properties, such as attractiveness, shared by neither extreme.

Since a long straight line, or linear array, of sets cannot be built up from shorter segments in a reliable way, for use as a definition of a linear array let us seek a reasonable general description of the conditions under which loops and detours can be avoided.

The Special Case of Nested Sets—The Guttman Scale

Consider a finite sequence of sets $S_1 , S_2 , \cdots , S_n$ such that $S_i \subset S_{i+1}$ for $i = 1, 2, \cdots , n$. These sets are nested within one another with S_n being the largest. Let such a sequence be called N.

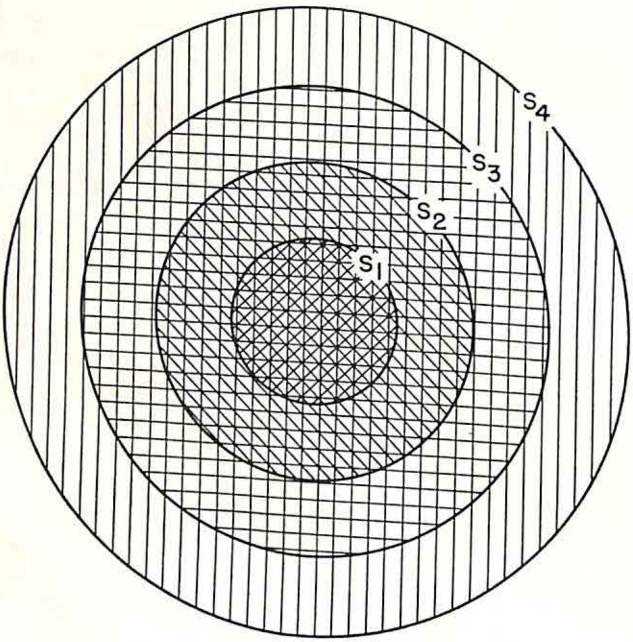


FIGURE 3

A nested sequence of sets as referred to in Theorem 4

THEOREM 4. *If S_i , S_j , and S_k are in N and $i \leq j \leq k$, then b_{ijk} .*

PROOF. $S_i \cap \bar{S}_j \cap S_k$ is a subset of $S_i \cap \bar{S}_j$ which is empty, and $\bar{S}_i \cap S_j \cap \bar{S}_k$ is a subset of $S_j \cap \bar{S}_k$ which is empty, establishing b_{ijk} . Q.E.D.

COROLLARY. *If* S_i, S_j, *and* S_k *are in* N *and* $i \leq j \leq k$, *then* $D_{ij} + D_{jk} = D_{ik}$.

At least up to this point, the material given is familiar and found in standard mathematical texts (e.g., [7], pp. 16, 24, 107, Exercise g).

The condition that an array of sets be nested is sufficient to ensure transitivity of betweenness and additivity of distances, but it seems quite restrictive in that it certainly does not correspond to qualitative scales. However, it is possible to make a qualitative or substitutive scale from two nested arrays and a constant set.

A Linear Array of Sets

The following development, like that above, will be restricted to finite arrays of sets. No great difficulties are foreseen in generalizing, but the ability to deal with finite numbers of sets is an advantage in empirical applications of the system.

DEFINITION 3. Let $A^* = \langle A_1, A_2, \cdots, A_n \rangle$ and $B^* = \langle B_1, B_2, \cdots, B_n \rangle$ be two nested n-tuples of sets, with the restriction that $A_n \cap B_n = \emptyset$ (whence all the A_i are disjunct from all the B_j). Let C be any set in $\mathbb{S}^*$ such that $A_n \cap C = B_n \cap C = \emptyset$. Then the n-tuple of sets

$$L^* = \langle L_1, L_2, \cdots, L_n \rangle,$$

where

$$L_i = A_i \cup B_{n-i+1} \cup C,$$

is a *linear array of sets.*

THEOREM 5. *If* L_i, L_j, *and* L_k *are in a linear array* L^*, *and* $i \leq j \leq k$, *then* b_{ijk}.

PROOF. Partition L_i into A_i, B_{n-i+1}, and C.

By the hypothesis that A^* is a nested sequence, and Theorem 4,

$$A_i \cap \bar{A}_j \cap A_k = \emptyset.$$

Similarly,

$$B_{n-i+1} \cap \bar{B}_{n-j+1} \cap B_{n-k+1} = \emptyset,$$

and certainly

$$C \cap \bar{C} \cap C = \emptyset.$$

Now,

$$L_i \cap \bar{L}_j \cap L_k = (A_i \cup B_{n-i+1} \cup C) \cap \overline{(A_j \cup B_{n-j+1} \cup C)} \cap (A_k \cup B_{n-k+1} \cup C)$$

which, because the A's, B's and C are all disjoint, is equal to

$$(A_i \cap \bar{A}_j \cap A_k) \cup (B_{n-i+1} \cap \bar{B}_{n-j+1} \cap B_{n-k+1}) \cup (C \cap \bar{C} \cap C) = \emptyset.$$

Similar argument shows that

$$\bar{L}_i \cap L_j \cap \bar{L}_k = \emptyset. \qquad \text{Q.E.D.}$$

COROLLARY. *If L^* is a linear array of sets and $i \leq j \leq k$, then*

$$D_{ij} + D_{jk} = D_{ik} .$$

PROOF. Theorems 5 and 2.

Note that nested sets are a special case of a linear array in which, for example, $B_i = \emptyset$ for $i = 1, 2, \cdots, n$. Therefore, Theorem 4 is a special case of Theorem 5.

The restrictions of a *linear array* (Definition 3) have been shown to be sufficient to ensure transitivity of betweenness. These same restrictions will now be shown to be a necessary condition for such transitivity.

THEOREM 6. *If $R^* = \langle R_1 , R_2 , \cdots , R_n \rangle$ is a sequence of sets such that for all $i, j, k, = 1, 2, \cdots , n$, and $i \leq j \leq k$, b_{ijk} , then R^* is a linear array of sets.*

PROOF. It must be shown that there exist nested sequences $A^* = \langle A_1 , \cdots , A_n \rangle$ and $B^* = \langle B_1 , \cdots , B_n \rangle$ and a set C such that $R_i = A_i \cup B_{n-i+1} \cup C$, and $A_n \cap B_n = \emptyset$, $A_n \cap C = \emptyset$ and $B_n \cap C = \emptyset$.

(i) Define, for an example,

$$C = R_1 \cap R_2 \cap \cdots \cap R_n \qquad (=R_1 \cap R_n),$$
$$A_i = R_i \cap \bar{R}_1 ,$$
$$B_j = R_{n-j+1} \cap \bar{R}_n .$$

(ii) Show that $A_i \subset A_{i+1}$. This is equivalent to the statement that

$$(R_i \cap \bar{R}_1) \subset (R_{i+1} \cap \bar{R}_1),$$

or

$$(R_i \cap \bar{R}_1) \cap \overline{(R_{i+1} \cap \bar{R}_1)} = \emptyset,$$

or

$$(R_i \cap \bar{R}_1) \cap (\bar{R}_{i+1} \cup R_1) = \emptyset.$$

Multiplying,

$$(R_i \cap \bar{R}_1 \cap \bar{R}_{i+1}) \cup (R_i \cap \bar{R}_1 \cap R_1) = \emptyset,$$

which is true: the first term is empty by $b_{1,i,i+1}$, and the second term is obviously empty. Similar steps show that the B's are nested.

(iii) Show that $A_i \cup B_{n-i+1} \cup C = R_i$.

$$\begin{aligned} A_i \cup B_{n-i+1} \cup C &= (R_i \cap \bar{R}_1) \cup (R_{n-(n-i+1)+1} \cap \bar{R}_n) \cup (R_1 \cap R_n) \\ &= (R_i \cap \bar{R}_1) \cup (R_i \cap \bar{R}_n) \cup (R_1 \cap R_n). \end{aligned}$$

Factoring the first two terms gives

$$\begin{aligned} A_i \cup B_{n-i+1} \cup C &= [R_i \cap (\bar{R}_1 \cup \bar{R}_n)] \cup (R_1 \cap R_n) \\ &= [R_i \cap (\overline{R_1 \cap R_n})] \cup (R_1 \cap R_n). \end{aligned}$$

By the hypothesis b_{1in}, it is seen that $R_1 \cap R_n = R_i \cap (R_1 \cap R_n)$. Substituting this in the above equation,

$$\begin{aligned} A_i \cup B_{n-i+1} \cup C &= [R_i \cap (\overline{R_1 \cap R_n})] \cup [R_i \cap (R_1 \cap R_n)] \\ &= R_i \ . \end{aligned}$$

The rest of the proof is trivial.

Theorem 6 shows, in a formal way, that any sequence of sets throughout which the relation of betweenness holds may be dissected into two nested sequences, running in opposite directions, and a constant remainder. This gives a sharp meaning to the concept of a substitutive scale, for as one moves from R_i to R_{i+1} , part of the elements of B_n are removed, and some new elements from A_n are substituted. There is no requirement that the measure added must equal the amount removed, though this may be an interesting special case.

From Theorem 6 it is seen that a linear array of sets is well represented by three generic sets, A^*, B^*, and C. A_n and B_n will be called the *poles*, the nested sequences A^* and B^* the *polar arrays*, and C the *core*, it being common to all elements in the original array.

Multidimensional Arrays

In current multidimensional scaling it is common to assume an arrangement in Euclidean space. Attneave [1] and Landahl [6] have suggested non-Euclidean rules for computing the distances between objects which vary on more than one dimension, suggesting that psychological intuition does not force us to the Euclidean assumption. The present discussion suggests a generalization of the above system to more than one dimension, giving rise to a non-Euclidean arrangement like those mentioned above.

DEFINITION 4. Two linear arrays of sets, L^* and L'^*, are said to be *parallel* if and only if their polar arrays are identical.

LEMMA 1. *If L^* and L'^* are parallel arrays,* $D_{ij} = D_{i'j'}$.

PROOF. This follows at once from the fact that distance depends only on the polar sets, not at all on the core (or common elements).

LEMMA 2. *If L^* and L'^* are parallel linear arrays,*

$$D_{L_i L_{j'}} = D_{ij'} = D_{ij} + D_{C.C'} .$$

PROOF. Partition the universe $\mathfrak{U}$ into the following cells:

$$Q_1 = L_i \cap L'_j \cap C \cap C'; \qquad Q_2 = L_i \cap L'_j \cap \bar{C} \cap \bar{C}';$$
$$Q_3 = \bar{L}_i \cap L'_j \cap \bar{C} \cap C'; \qquad Q_4 = L_i \cap \bar{L}'_j \cap C \cap \bar{C}';$$
$$Q_5 = L_i \cap \bar{L}'_j \cap \bar{C} \cap \bar{C}'; \qquad Q_6 = \bar{L}_i \cap L'_j \cap \bar{C} \cap \bar{C}';$$
$$Q_7 = \bar{L}_i \cap \bar{L}'_j \cap \bar{C} \cap \bar{C}'.$$

The cells are clearly mutually exclusive. They are exhaustive because seven other possible cells are subsets of $C \cap \bar{L}_i$ (empty) or $C' \cap \bar{L}'_j$ (also empty), and the other two possible cells, $L_i \cap L'_j \cap \bar{C} \cap C'$ and $L_i \cap L'_j \cap C \cap \bar{C}'$ are identical with Q_2, being comprised of the elements common to L_i and L'_j which are not in the cores. A diagram of the sets and the partition is shown in Figure 4. Here, $D_{ij} = m(Q_5) + m(Q_6)$, $D_{C.C'} = m(Q_3) + m(Q_4)$,

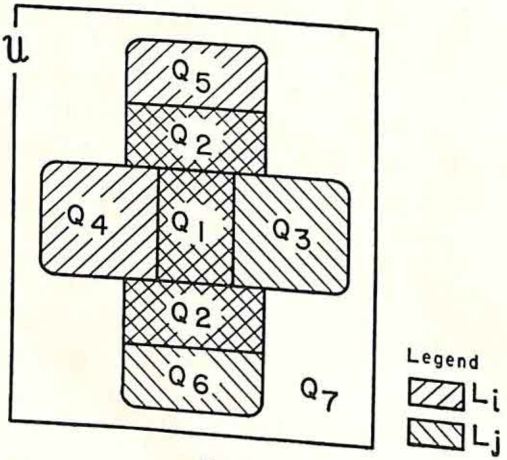


FIGURE 4

The sets L_i and L'_j and the partition referred to in Lemma 2

and as stated by the lemma, $D_{ij'} = m(Q_3) + m(Q_4) + m(Q_5) + m(Q_6) = D_{ij} + D_{CC'}$. Q.E.D.

In what follows, a set of parallel linear arrays will be referred to as a *dimension*, using the following notation. A dimension X_L^* is a set of linear arrays, $L^{*(1)}, L^{*(2)}, \cdots, L^{*(t)}$, which are parallel to one another.

DEFINITION 5. If $X_L^* = L^{*(1)}, \cdots, L^{*(t)}$ is a dimension and the cores of the arrays, $C^{(1)}, \cdots, C^{(t)}$ form a linear array, then the set of members of $L^{*(1)}, \cdots, L^{*(t)}$ is called a *linear two-space.*

Theorem 7. *If S^* is a linear two-space made up of the linear arrays*

$$\begin{array}{lcl} L^{*(1)} & = & \langle L_1^{(1)}, \cdots, L_r^{(1)}, \cdots, L_n^{(1)} \rangle, \\ \vdots & & \quad\vdots \qquad\quad \vdots \qquad\quad \vdots \\ L^{*(t)} & = & \langle L_1^{(t)}, \cdots, L_r^{(t)}, \cdots, L_n^{(t)} \rangle, \end{array}$$

then the sequence of sets composed of the rth *member of each array, i.e.,* $L_r^{(1)}, \cdots, L_r^{(i)}, \cdots, L_r^{(t)}$ *is also a linear array.*

Proof. By definition of a two-space the cores form a linear array, and by adding to each core the constant set $L_r^{(i)} \cap \overline{C^{(i)}}$, the result is still a linear array. Q.E.D.

Theorem 7 shows that there is no asymmetry between the two dimensions of a two-space. If X^* and Y^* are the two dimensions, the cores of X^* give the polar arrays of Y^*, and the cores of Y^* give the polar arrays of X^*.

Further dimensions may be constructed by considering the joint core of a two-space, i.e., the intersection of all sets in the space. If one were to construct a sequence of parallel two-spaces, identical except for their joint cores, and if the resulting sequence of joint cores were a linear array, then one would have a three-space. This process can be repeated indefinitely, so that an n-space is defined for any n.

A Non-Pythagorean Theorem, and Description of the Space

An advantage of Euclidean space, which has contributed to its use in psychological measurement, is the Pythagorean theorem, that the distance between two points which are apart, on various dimensions, by amounts $x_1, x_2, \cdots, x_n$, is given by $d = \sqrt{\sum_n x_i^2}$. If two sets in an n-space of the present type are apart, on the various dimensions, by amounts $x_1, x_2, \cdots, x_n$, then the distance between them is $d = \sum_n x_i$. This is shown informally by applying Lemmas 1 and 2. The only differential elements must, by definition of an n-space, be elements in the linear arrays in the n dimensions. Since the poles of the various dimensions are discrete, the measures are all added. That the result is unique follows from Lemma 2 and the symmetry of dimensions.

One gets an idea of the space by considering a small finite two-space, rectangular in its Euclidean mapping, like that in Figure 5. The sets are represented by points, and the *paths* from point a to point b are indicated.

The number of paths from a to b in any two-space is easily computed. If a and b are separated by r sets in one dimension and by s sets in the other, the number of paths is $\binom{r+s}{r}$. Of course, the steps need not be all of the same distance, but the total lengths of all paths will be the same.

Along with the concepts of parallel arrays, and perpendicular arrays (in which the poles have no overlap), one can also consider oblique arrays (in which there is some overlap between the poles). In Figure 5, with the particular space of sets available, one proceeds from a to b by changes in one dimension at a time. It would be possible, if sets can be produced experimentally (as where they are physical objects in a perception experiment) to introduce new sets whereby one goes directly from a to b by successive changes in *both* dimensions simultaneously. The total distance will be the same, of course, because one either introduces more intermediate sets or has

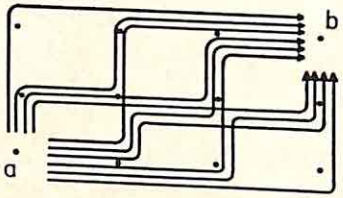


FIGURE 5
A two-space with 3 x 4 sets, each mapped on a point
(The ten ways of going from a to b are indicated.)

sets which, because they differ in two dimensions, are farther apart than the ones shown in Figure 5. However, this possibility shows that the present geometry does not permit formal identification of *pure variables*. If one had two pure psychological variables, each of which gives rise to linear arrays, then a new linear array could be made by uniting the first two, element by element (provided, of course, that they are of the same number of elements). This new compound array would have all the formal properties of a linear array and would be indistinguishable from a pure array in this system. The same fact applies to Guttman scales as a special case and is presumably rather general, suggesting that a pure variable cannot be isolated by internal evidence within this system.

Discussion

The similarity between the axioms of a measure function and the metric axioms has been used to devise a way of measuring the distance between sets. Hays [5] has used the same distance concept, calling it "implicational difference," and defining it by

$$\mathrm{Di}_{AB} = m(A) + m(B) - 2m(A \cap B),$$

where A and B are sets. In this paper distance is defined as the measure of the symmetric set difference, which is identical with Hays' formulation.

Certain related concepts of the distance between sets may be mentioned for contrast. Galanter [4] bases his distance measure on the ratio of the measure of the set difference to the measure of the union of the two sets. If this distance is called G_{AB}, it would be

$$G_{AB} = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B) - m(A \cap B)}.$$

This distance is restricted between zero and one and does not have the desirable property of additivity when sets are ordered. Bush and Mosteller [3] define a similarity index η as the ratio of the measure of the intersection to the measure of one of the sets. That is,

$$\eta(A \text{ to } B) = m(A \cap B)/m(A).$$

This index cannot be the basis of a distance since it is not generally symmetric. If one used the mean or sum of the two directional η's, the result would be similar to Galanter's proposal.

With respect to the ordering of sets, both Galanter [4] and Burke [2] have presented ideas like the present one, and the writer also previously proposed a similar ordering [8]. Burke distinguished between substitutive and nested orderings in essentially the way done here, but did not give a general treatment of substitutive orderings and ways of escaping the intransitivity of the local relationship. Hays [5] does not consider an ordering of the present type, and instead embeds the numerical distances in a Euclidean space. In the writer's opinion, this loses part of the advantage of the set-theoretic formulation.

There are occasions in psychological theory in which a dimensional or metric approach is entirely natural and valuable, as in dealing with loudness, pitch, or hue, in the discussion of generalization gradients in learning, in treating polarized attitudes, and in certain applications of statistics. On the other hand, the categorical approach has a firmer logical foundation and the advantage that raw data in psychological experiments are usually categorical. The present paper is an attempt to bridge the gap between the two approaches by showing a way to develop dimension from purely set-theoretic concepts. It is hoped that this may help in unifying the mathematical approaches to psychological problems.

REFERENCES

[1] Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516-556.

[2] Burke, C. J. Some problems in discrimination models. *Amer. Psychologist*, 1956, **11**, 483.

[3] Bush, R. R. and Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413-423.

[4] Galanter, E. H. An axiomatic and experimental study of sensory order and measure. *Psychol. Rev.*, 1956, **63**, 16-28.

[5] Hays, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri and L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univ. Press, 1958.

[6] Landahl, H. D. Neural mechanisms for the concepts of difference and similarity. *Bull. math. Biophysics*, 1945, **7**, 83-88.

[7] Munroe, M. E. *Introduction to measure and integration.* Cambridge, Mass.: Addison Wesley, 1953.

[8] Restle, F. A theory of discrimination. Unpublished doctoral dissertation, Stanford Univ., 1953.

Manuscript received 10/2/58

Revised manuscript received 3/2/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS USING THE LOGISTIC FUNCTION

A. E. MAXWELL

INSTITUTE OF PSYCHIATRY, LONDON UNIVERSITY

The logistic function is proposed as an alternative to the integrated normal function when estimating parameters of test items. The logistic curve is described; an iterative method for finding maximum likelihood estimates of its parameters is given, and an example of its use is presented.

Finney [9] pointed out that the problem of estimating the *limina* and *precisions* of test items, discussed earlier by Ferguson [8], was closely analogous to problems in toxicological experiments. In particular he showed that the methods of probit analysis could be applied with advantage to Ferguson's data to obtain efficient estimates of the parameters in question. More recently Berkson [2, 3] and Anscombe [1] have shown that in toxicological work a useful alternative to the probit model, in which an integrated normal response law is assumed, is one in which the logistic function is employed. They point out that while the shapes of these two types of response curve are very similar, simple sufficient statistics are available for the parameters of the logistic curve. Attracted by this property Birnbaum [5], while "considering the application of the Neyman-Pearson and Wald theories of inference and statistical decision making to problems of efficient design and use of tests of a single ability," demonstrated some advantages of adopting the logistic function, rather than the usual normal ogive, for ascertaining the probability that an examinee of given ability would have a specific response pattern to a k-item test. While Birnbaum's approach (also illustrated in [6] and [7]) should be viewed as a portent of the shape of future test theory, it is useful meanwhile to look at the logistic model as an alternative to the probit model for the straightforward estimation of the limina and precision of test items.

The Problem as Conventionally Stated

Conventionally (for references see [11]) it has been assumed that the probability P_{ik} that an examinee k, of "true" ability x_k on a test sampling an ability x, will answer the item j of the test correctly is given by the expression

$$P_{ik} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x_k - \alpha_j)/\sigma_j} e^{-u^2/2}\, du, \tag{1}$$

where α_j and σ_j are constants for that item. When P_{jk} is 0.5, x_k is equal to α_j so that, following psychophysical nomenclature, α_j is the *limen* of the item, that point on the x-scale below which 50 percent of the population fail the item. The constant σ_j determines how well the item discriminates between individuals of high and low ability. The problem is one of finding efficient estimates, $\hat{\alpha}_j$ and $\hat{\sigma}_j$, of the population parameters α_j and σ_j. For completeness special reference should be made here to Lord [11], where he points out that (1) can be obtained, if desired, from certain assumptions, prominent amongst which are

(*i*) that there is a continuous variable x underlying the item, and
(*ii*) that x is normally distributed.

Lord then shows that if the metric x is the common factor of a set of items, measuring a single ability, x_k is invariant no matter what test of the ability is administered, and that this invariance holds whether or not x is normally distributed in the group tested. It should be noted that the common factor referred to here is not determined by a rank-one matrix of interitem tetrachorics (except in the special case where the common factor is normally distributed). It is found by an analog of latent structure analysis.

The probit of the proportion P_{jk} is defined [9] as the value of Y satisfying the equation

$$P_{jk} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(Y-5)} e^{-u^2/2}\, du, \tag{2}$$

where Y is the unit normal deviate corresponding to P_{jk} increased by 5. The probit of P is then linearly related to x by the equation

$$Y = 5 + (x - \alpha_j)/\sigma_j , \tag{3}$$

and by the familiar methods of probit analysis the constants α_j and σ_j can be estimated.

The Logistic Model

Suppose that groups of individuals are tested at each of k different values of the variable x_i ($i = 1, 2, \cdots, k$). In toxicological work x_i would stand for the ith dose of bactericide, but for present purposes it stands for the midpoint of the ith interval in a frequency distribution of test scores. Of the n_i individuals falling in this interval, r_i, or a proportion $p_i = r_i/n_i$, are found to have passed the jth item of the test. It is now assumed that the p_i have independent binomial distributions, that $P_i = 1 - Q_i$ is the proportion of individuals in the population falling in the interval x_i who pass the jth item, and that P_i is a given function of two parameters α and β of the item concerned. The logistic law states that

$$P_i = [1 + \exp(-\alpha - \beta x_i)]^{-1}. \tag{4}$$

The logit of the proportion P_i, generally denoted by L_i, is defined as

$$L_i = \ln (P_i/Q_i) = \alpha + \beta x_i . \tag{5}$$

Writing $\alpha + \beta x_i = \beta(x_i - \mu)$, μ, which is equal to $-\alpha/\beta$ corresponds to $P = 0.5$, so that $-\alpha/\beta$ is the limen of the item.

For each of (2) and (4) it is seen that a transformation of the dependent variable P can be made such that if the transformed value is plotted against x a linear relationship results, the slope and intercept of which give the parameters of the original function.

The Shape of the Logistic Function

If x in (4) ranges from $-\infty$ to $+\infty$, P ranges from 0 to 1. Differentiating (4) with respect to x,

$$dP/dx = \beta P(1 - P). \tag{6}$$

Hence $dP/dx = 0$, when $P = 0$ and 1, so that the logistic curve has both a lower and an upper asymptote.

Differentiating (6),

$$d^2P/dx^2 = (1 - 2P), \tag{7}$$

which shows that the curve has a point of inflection at -0.5, halfway between the two asymptotes. The abscissa of this point, as already noted, is $-\alpha/\beta$, the limen of the item. From this description it is seen that the logistic curve resembles very closely the normal ogive. For this reason, it is, on purely empirical grounds, suitable for fitting to proportions of individuals of different levels of ability passing a test item [8].

Maximum Likelihood Estimates of the Logistic Parameters

In the ith interval, the probability that p_i is the proportion passing the item is

$$\frac{n_i\,!}{r_i\,!\,(n_i - r_i)!} P_i^{r_i} Q_i^{n_i - r_i}. \tag{8}$$

The likelihood function over all intervals is

$$\prod_i \frac{n_i\,!}{r_i\,!\,(n_i - r_i)!} P_i^{r_i} Q_i^{n_i - r_i}. \tag{9}$$

The terms in the logarithm of this function which involve the unknown parameters, α and β, are given by

$$L = \sum_i \{r_i \log_e P_i + (n_i - r_i) \log_e Q_i\}. \tag{10}$$

Substituting in (10) the value of P_i from (4) and differentiating with respect to α and β, the estimating equations for the two parameters are found to be

$$\sum n_i(p_i - \hat{P}_i) = 0, \tag{11}$$

and

$$\sum n_i x_i(p_i - \hat{P}_i) = 0, \tag{12}$$

respectively, where $\hat{P}_i = [1 + \exp(-\hat{\alpha} - \hat{\beta}x_i)]^{-1}$, $\hat{P}_i$, $\hat{\alpha}$ and $\hat{\beta}$ being estimates of the population parameters P_i, α, and β. These equations, based on the sufficient statistics $\sum n_i p_i$ and $\sum n_i p_i x_i$, have been given by Berkson [2, 4], who also shows how to solve them iteratively. The procedure is as follows. Given trial values α_0 and β_0 of the parameters required, a trial value l_0 of the logit is obtained using (5); it is $l_0 = \alpha_0 + \beta_0 x_i$, and corresponds to an estimate $\hat{p}_0$ of P_i given by

$$\hat{p}_0 = [1 + \exp(-l_0)]^{-1}.$$

Using this result, and retaining only the first term in a Taylor expansion,

$$\hat{p}_0 - \hat{P}_i \simeq (\hat{l}_0 - \hat{l}_i)\hat{P}_0\hat{Q}_0 .$$

Substituting this approximation in (11) and (12), it will be seen that the differentials $\delta\hat{\alpha}$ and $\delta\hat{\beta}$, which are the corrections to the trial values of the parameters, are given by the equations

$$\begin{aligned} \sum n_i p_i - \sum n_i \hat{p}_0 &= \delta\hat{\alpha}\left(\sum w_i\right) + \delta\hat{\beta}\left(\sum w_i x_i\right), \\ \sum n_i p_i x_i - \sum n_i \hat{p}_0 x_i &= \delta\hat{\alpha}\left(\sum w_i x_i\right) + \delta\hat{\beta}\left(\sum w_i x_i^2\right), \end{aligned} \tag{13}$$

where $w_i = n_i p_0 q_0$, corresponding to x_i. Solving these equations the corrections are found to be

$$\delta\hat{\beta} = \frac{\sum n_i p_i x_i - \sum n_i \hat{p}_0 x_i - \sum w_i x_i\left(\sum n_i p_i - \sum n_i \hat{p}_0\right)/\sum w_i}{\sum w_i x_i^2 - \left(\sum w_i x_i\right)^2/\sum w_i} \tag{14}$$

and

$$\delta\hat{\alpha} = \frac{\sum n_i p_i - \sum n_i \hat{p}_0 - \delta\hat{\beta} \sum w_i x_i}{\sum w_i}. \tag{15}$$

Anscombe [1] points out that it is unnecessary in further iterations to recalculate the coefficients on the right-hand side of equations (13), and since the expressions $\sum n_i p_i$ and $\sum n_i p_i x_i$ remain unchanged from iteration to iteration only the terms $\sum n_i \hat{p}_0$ and $\sum n_i \hat{p}_0 x_i$ have to be recalculated. "This procedure," Anscombe ([1], p. 461) remarks, "is to be compared with Finney's which is exactly modelled on the ingenious procedure due to Fisher and Bliss for fitting the integrated-normal law, for which no such sufficient statistics are available."

Estimates of the variance of $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\mu}$ are given [3] by the formulas

$$\begin{aligned}
&\text{var}\ \ (\hat{\beta}) = 1/\sum w_i(x_i - x)^2; \\
(16)\qquad &\text{var}\ \ (\hat{\alpha}) = (1/\sum w_i) + \bar{x}^2\ \text{var}\ (\hat{\beta}); \\
&\text{var}\ \ (\hat{\mu}) = [\text{var}\ (\hat{\alpha}) + \text{var}\ (\hat{\beta})(\hat{\mu} - x)^2]/\hat{\beta}^2;
\end{aligned}$$

where $\bar{x} = \sum w_i x_i / \sum w_i$.

A quick graphical method for estimating α and β has recently been provided by Hodges [10].

An Example

The data used in the example are taken from Ferguson's article [8] and are the same data as employed by Finney [9] in his example using probit analysis. Omitting the two extreme values of x, where the proportions may be unreliable, estimates α_0 and β_0 of the parameters were read from a rough plot of the variate x_i as abscissa and $l_i = \log_e p_i/q_i$, where $q_i = 1 - p_i$, as ordinate. These were 2.0 and 1.4 respectively, giving the equation $l_0 = 1.4\,x + 2.0$. Values of l_0 for each x-value were now calculated; these were converted to p_0-values using the table prepared by Berkson [4], and from the same table the weights p_0q_0 were read. These data are tabulated in Table 2, while below the table the quantities required for substitution in (14) and (15) are recorded.

TABLE 1

Proportions (p_i) Passing an Item at Different Levels of Ability (x_i)

x_i	n_i	p_i	n_ip_i
-1.94	15	0.33	4.95
-1.16	25	0.56	14.00
-0.58	43	0.70	30.10
0.00	50	0.94	47.00
0.58	43	0.95	40.85
1.16	25	0.96	24.00
1.94	15	0.93	13.95

$\sum n_ip_i = 174.85$ $\qquad$ $\sum n_ip_ix_i = 35.2950$

TABLE 2

Data for the First Iteration

x_i	$\hat{\ell}_o$	$\hat{p}_o$	$\hat{p}_o\hat{q}_o$	$n_i\hat{p}_o$	$w_i = n_i\hat{p}_o\hat{q}_o$
-1.94	-0.72	0.32739	0.22021	4.91085	3.30315
-1.16	0.38	0.59387	0.24119	14.84675	6.02975
-0.58	1.19	0.76674	0.17885	32.96982	7.69055
0.00	2.00	0.88080	0.10499	44.04000	5.24950
0.58	2.81	0.94321	0.05356	40.55803	2.30308
1.16	3.62	0.97392	0.02540	24.34800	0.63500
1.94	4.72	0.99116	0.00876	14.86740	0.13140

$\sum w_i = 25.34243$

$\sum w_i x_i = -15.535838$

$\sum w_i x_i^2 = 25.25622$

$(\sum w_i x_i)^2 = 241.36226$

$\sum n_i \hat{p}_o = 17654085$

$\sum n_i \hat{p}_o x_i = 34,73832$

$\sum n_i n_i - \sum n_i \hat{p}_o = -1.69085$

$\sum n_i p_i x_i - \sum n_i \hat{p}_o x_i = 0.55668$

Substituting the results from Tables 1 and 2 in (14) and (15), the first corrections are $\delta\hat{\beta} = -0.030502$ and $\delta\hat{\alpha} = -0.085419$ so that first approximations to the maximum likelihood estimates of the parameters are $\hat{\beta} = 1.3695$ and $\hat{\alpha} = 1.9146$; from these the limen of the item is found to be $-1.9146/1.3695 = -1.3980$.

For the second iteration the logit equation is $\hat{l}_0 = 1.3695x + 1.9146$; one now recalculates values for the expressions $\sum n_i\hat{p}_0$ and $\sum n_i\hat{p}_0 x_i$. These are 174.79230 and 35.26088, respectively, and the corrections in the second iteration are found to be $\delta\hat{\beta} = 0.004416$ and $\delta\hat{\alpha} = 0.004975$. Since these corrections have nonzero entries only after the second decimal places, further iterations are unnecessary for present purposes. The adjusted values of $\hat{\beta}$ and $\hat{\alpha}$ now are 1.3739 and 1.9196, respectively, while the new estimate of the limen of the item is -1.3972. The three successive estimates of the limen of the item, using the logistic function, are therefore -1.4286, -1.3980, and -1.3972. Convergence seems to be good.

Corresponding to the first two of these, Finney, using the probit method, obtained the values -1.32 and -1.471, respectively. While he considered his fit after one iteration to be sufficiently accurate it appears that it would have been preferable, both from the point of view of demonstrating convergence and in order to get a more accurate answer, had he performed at least one further iteration.

Using (16), estimates of the standard errors of $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\mu}$ are found to be

$$\text{S.E. of } \hat{\beta} = 0.0540, \quad \text{S.E. of } \hat{\alpha} = 0.2014, \quad \text{S.E. of } \hat{\mu} = 0.1956.$$

The estimate 1.3739 of β, the gradient of the logit line, is a measure of the discrimination value of the item. When compared with 0.0540, the estimate of its standard error, it is found to be highly significant.

REFERENCES

[1] Anscombe, F. J. On estimating binomial response relations. *Biometrika*, 1956, **43**, 461-464.

[2] Berkson, J. Why I prefer logits to probits. *Biometrics*, 1951, **7**, 327-339.

[3] Berkson, J. A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. statist. Ass.*, 1953, **48**, 565-599.

[4] Berkson, J. Tables for the maximum likelihood estimate of the logistic function. *Biometrics*, 1957, **13**, 28-34.

[5] Birnbaum, A. Probability and statistics in item analysis and classification problems. Efficient design and use of tests of mental ability for various decision-making problems. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[6] Birnbaum, A. Probability and statistics in item analysis and classification problems. Further considerations of efficiency in tests of a mental ability. Tech. Rep. No. 17, Project No. 7755-23, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[7] Birnbaum, A. Probability and statistics in item analysis and classification problems. On the estimation of mental ability. Series Rep. No. 15, Project No. 7755-523, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[8] Ferguson, G. A. Item selection by the constant process. *Psychometrika*, 1942, **7**, 19-29.

[9] Finney, D. J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, **9**, 31-39.

[10] Hodges, J. L., Jr. Fitting the logistic by maximum likelihood. *Biometrics*, 1958, **14**, 453-461.

[11] Lord, F. M. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 1953, **18**, 57-76.

Manuscript received 6/19/58

Revised manuscript received 2/21/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

THREE MULTIVARIATE MODELS: FACTOR ANALYSIS, LATENT STRUCTURE ANALYSIS, AND LATENT PROFILE ANALYSIS[*]

W. A. GIBSON[†]

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE

The factor analysis model and Lazarsfeld's latent structure scheme for analyzing dichotomous attributes are derived to show how the latter model avoids three knotty problems in factor analysis: communality estimation, rotation, and curvilinearity. Then the latent structure model is generalized into latent profile analysis for the study of interrelations among quantitative measures. Four latent profile examples are presented and discussed in terms of their limitations and the problems of latent metric and dimensionality thereby raised. The possibility of treating higher order empirical relations in a manner paralleling their various uses in the latent structure model is indicated.

At an early point in *Multiple-Factor Analysis* ([18], p. 70), Thurstone remarks:

> It would be unfortunate if some initial success with the analytical methods to be described here should lead us to commit ourselves to them with such force of habit as to disregard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

This paper is an attempt to take that statement to heart.

First the derivation of the factor analysis model will be sketched, noting three inherent conceptual and procedural problems: (i) how to estimate communalities in the event that only shared variance is to be analyzed, (ii) how to resolve rotational indeterminacy, and (iii) what to do with the extra linear factors that are forced to emerge when nonlinearities occur in the data. Some recent concepts for the multidimensional analysis of qualitative data—the latent structure model of Lazarsfeld [15]—are considered, with special reference to their handling of the trouble spots in factor analysis. Next these new concepts are generalized to produce an alternative way of analyzing the interrelations among quantitative measures. This is the latent profile model.

All three of these models are discussed strictly from the point of view of sample statistics. The problem of generalizing to a population of which the sample may be considered representative is not taken up for any of the three models.

---

[*] The latter model is anticipated in an earlier paper by Green [12].

[†] The major portion of this paper was completed at the Center for Advanced Study in the Behavioral Sciences. The opinions expressed are those of the author and are not to be construed as reflecting official Department of the Army policy.

Four examples of the latent profile model are given. Two of them are fictitious but their form closely parallels the corresponding two empirical examples. These examples are discussed from the viewpoint of the further work they suggest concerning the metric and dimensionality of the latent space. The use of higher order empirical relations for testing fit and for further particularization of the latent profile model is briefly discussed, and the possibility of close parallelism between developments in the latent structure and latent profile models is pointed out. The paper concludes with a plea for continued flexibility in the choice of multivariate models as new and improved ones appear.

Some Problems in Factor Analysis

The fundamental postulate of factor analysis ([18], p. 63) is expressed in the simple linear equation

$$Z_{ij} = a_{j1}Z_{i1} + a_{j2}Z_{i2} + \cdots + a_{jq}Z_{iq} . \tag{1}$$

Z_{ij} is the standard score of individual i on test j. Z_{i1} , Z_{i2} , $\cdots$, Z_{iq} are the standard scores of individual i on a hypothesized set of q statistically independent traits or factors. (This does not preclude subsequent conversion to correlated factors in any given analysis. The algebra of correlated factors will not be introduced here, however, for that would only complicate the discussion without changing the arguments.) The a's in (1) are a set of weights descriptive of test j and invariant for individuals.

Straightforward summational algebra and the independence of factors lead directly from (1) to the basic equation of factor analysis ([18], p. 78):

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jq}a_{kq} . \tag{2}$$

Thus r_{jk} , the correlation between tests j and k, is expressed as a simple bilinear function of the a's for those two tests. These a's, also known as factor loadings, are interpretable as correlations between tests and factors.

The essential task in the factor analysis of a battery of s tests is to solve for the a's in the system of $s(s - 1)/2$ bilinear equations resulting from (2). The number of factor loadings is sq, q for each of the s tests. These loadings can be obtained by a wide variety of techniques that have been developed over the years. Most of these methods attempt to account for the intercorrelations in terms of a minimum number of factors. Perfect accounting for the intercorrelations by the factors is seldom demanded because of sampling error and the frank expectation of at least minor disagreements between model and data. Some nonvanishing "residual" correlations are permitted, so long as they are small and show no systematic pattern.

One troublesome feature of the factor model is the problem of how to deal with those elements in the correlation matrix that have repeated subscripts—the diagonal r_{jj}'s. To take these as perfect self-correlations of unity would amount to trying to analyze all of the variance of every test, including the unreliable part. To insert the test reliabilities into the diagonal cells would imply an interest in analyzing all of the "true" or repeatable variance of the tests, including that specific to each test and unrelated to other tests in the battery. More commonly preferred among factor analysts is to attempt to analyze only that part of a test's variance—its communality—that is shared with other tests in the battery. Communalities could be defined alternatively as those portions of the self-correlations accounted for by the factors that suffice to account for the between-test correlations. In any event, the communalities are not empirically given and yet are needed at the start for maximum efficiency of solution. In principle the communalities could be determined by certain operations (cf. [18], pp. 294–307) applied to the empirically given side entries in the correlation matrix, but these operations are usually so time consuming that a successive approximations approach is often substituted. Rough communality estimates are used to obtain an initial factorial solution, which in turn provides improved estimates for a second cycle of the same kind, and so on until the communalities are sufficiently "stabilized." It fortunately happens that with large test batteries little or no iterating of this kind is needed, but for small batteries several cycles may be required before the communality problem is adequately resolved.

A second and more important problem in factor analysis is the inherent partial indeterminacy of the a's that is known as the rotational problem. It is most easily understood in terms of q-dimensional geometry. The a's for any test j may be thought of as the projections of a point j on a set of coordinate axes in q-space. The table of factor loadings for the s tests then defines a configuration of s points in terms of their projections on a q-dimensional reference frame. But the equations of factor analysis, by themselves, do not indicate *which* position of the reference frame, among an infinite number of possibilities in the same q-space, is to be preferred. Only the origin of the coordinate system is fixed, so that the reference frame can be rotated freely from any position to any other without distortion of the configuration of points defined by it. Naturally the a's change with such rotations, but always in such a way as to preserve the spatial interrelations among the points they define. Many ways of resolving this rotational indeterminacy have been proposed. Probably the most notable is the simple structure principle ([18], pp. 181–193), which strives to simplify the factorial structure of the tests by maximizing the number of near-zero factor loadings. Many of these proposals (especially those centering around the simple structure principle) involve heavy computing loads, relatively rare geometric intuition, or both; many are debated or debatable; all are in the nature of afterthoughts that are not built into the equations defining the model.

Another perplexing problem in the factor analysis model is the paradox of difficulty factors [2, 4, 13, 20]. If factor analysis is applied to a battery of tests varying widely in difficulty but quite obviously measuring but one underlying trait, the result is not one but several "factors"—one for each level of difficulty. This is generally attributed to curvilinear relations among tests markedly different in difficulty, such curvilinearity being forced by the differential skewness of the score distributions. Note, however, that it is only the relations between tests and factors, as indicated by (1), that the factor model explicitly restricts to linear form.

The record of empirical fruitfulness (or lack thereof) of factor analysis is not at issue here. Nor are misapplications of it pertinent to this discussion. The other two models to be discussed here are meant to be put to much the same use, and they may very well suffer the same kind of misuse.

Some New Concepts: Latent Structure Analysis

Only one variety of latent structure analysis—that known as the discrete class model—is discussed here.

Latent structure analysis [15] is Lazarsfeld's technique for analyzing the interrelations among dichotomous attributes, such as the item responses on a survey questionnaire. It is based on linear recruitment equations of the following kind:

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q \,, \\
n_i &= n_1 p_{1i} + n_2 p_{2i} + \cdots + n_q p_{qi} \,, \\
n_{jk} &= n_1 p_{1jk} + n_2 p_{2jk} + \cdots + n_q p_{qjk} \,, \\
n_{jkl} &= n_1 p_{1jkl} + n_2 p_{2jkl} + \cdots + n_q p_{qjkl} \,, \quad \text{etc.}
\end{aligned} \tag{3}
$$

The quantities on the left are empirically given or *manifest* data. They indicate the number of people in the entire sample, n, the number endorsing a single item, n_i , the number endorsing any pair of items, n_{jk} , and so on. The quantities on the right are the underlying or *latent* parameters of the model. The number of terms on the right is q, the number of mutually exclusive and exhaustive subgroups (*latent classes*) into which the analysis will divide the total sample. The number of people in latent class 1 is n_1 , and so on. The *latent probability* p_{1j} is the proportion of the members of latent class 1 who endorse item j, p_{1jk} is the proportion of class 1 members who endorse both items j and k, and so on. Equations (3) merely show how the manifest joint endorsements are recruited from the latent classes. It may be noted, in passing, that equations (3), being a set of recruitment equations, are intrinsically linear, while the initial equation of factor analysis is linear only because it was made so.

The manifest-latent distinction that is so prominent in latent structure theory is of course central in factor analysis as well. There the tests and their intercorrelations are manifest, and the factors and the loadings thereon are latent. The same distinction also appears in the latent profile model to be discussed later.

The preceding recruitment equations are equally valid for any method of classification and for any number of latent classes. The next step is to invoke a pertinent basis for classification. This is the core of the latent structure model. It is quite relevant to require that each latent class exhibit homogeneity with respect to any underlying dimensions that may be responsible for the manifest interrelations. Perfect homogeneity is not crucial, so long as deviations from the class norm are random. Such random deviations are of course uncorrelated within the class. Thus it is sufficient to require that each latent class be homogeneous enough, with respect to any and all such latent dimensions, so that all item responses *within the class* are independent in the coin-tossing sense. This intraclass independence is expressed in the following equations.

$$
\begin{aligned}
&p_{1jk} = p_{1j}p_{1k}\,, \qquad p_{2jk} = p_{2j}p_{2k}\,, \cdots, p_{qjk} = p_{qj}p_{qk}\,,\\
&p_{1jkl} = p_{1j}p_{1k}p_{1l}\,, \qquad p_{2jkl} = p_{2j}p_{2k}p_{2l}\,, \cdots, p_{qjkl} = p_{qj}p_{qk}p_{ql}\,, \quad \text{etc.}
\end{aligned}
\tag{4}
$$

The substitution of (4) into (3) yields the basic equations of latent structure analysis ([15], p. 385).

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q\,,\\
n_j &= n_1p_{1j} + n_2p_{2j} + \cdots + n_qp_{qj}\,,\\
n_{jk} &= n_1p_{1j}p_{1k} + n_2p_{2j}p_{2k} + \cdots + n_qp_{qj}p_{qk}\,,\\
n_{jkl} &= n_1p_{1j}p_{1k}p_{1l} + n_2p_{2j}p_{2k}p_{2l} + \cdots + n_qp_{qj}p_{qk}p_{ql}\,, \quad \text{etc.}
\end{aligned}
\tag{5}
$$

Thus all of the manifest joint frequencies are accounted for in terms of $(q + sq)$ latent parameters, q class sizes and q latent probabilities $(p_{1j}, p_{2j}, \cdots, p_{qj})$ for each of the s items. The successive levels of manifest frequencies $(n, n_j, n_{jk}$, etc.) number, respectively, 1, s, $s(s-1)/2$, etc., the coefficients in the binomial expansion $(a + b)^s$. These add up to 2^s, the number of equations relating manifest to latent data in this model.

The task here, as in factor analysis, is to solve the basic equations for the unknown latent parameters. Several latent structure solutions already exist. The most recent of these [1, 7] avoid the use of any joint frequencies with repeated subscripts $(n_{jj}, n_{jjk}, n_{jjj}, n_{jjkl}$, etc.). In latent structure analysis these are treated as analogous to the communalities of factor analysis, in not being manifest. To interpret them as equivalent to the corresponding lower order joint frequencies without repeated subscripts (i.e., $n_{ii} = n_i$, $n_{ijk} = n_{jk}$, $n_{iii} = n_i$, etc.) would be analogous to the use of unit self-correlations in factor analysis. Instead, the latent structure model usually treats these elements with repeated subscripts as stemming from, rather than leading to, the latent parameters of the model that suffices to account for the manifest data *without* repeated subscripts.

The Anderson latent structure solution and an earlier one by Green [11] eliminate the latent structure analogue of the rotational problem through the use of manifest data with more than two subscripts, such as n_{jkl}. (A cursory investigation indicates that it might be self-contradictory for the factor model to make use of manifest interrelations among more than two variables at a time. This is because at least some of the joint distributions among factor scores would have to exhibit asymmetries that could easily destroy the basic linear postulate of factor analysis.)

These higher order data select, from the infinite number of rotational solutions accounting for lower order manifest frequencies equally well, the *one* that fits *themselves* best. Another early latent structure solution [5, 6, or 8] effects a partial (often severe) reduction of rotational indeterminacy without the expense of obtaining higher order data. This is accomplished by capitalizing on the simple fact that the latent parameters, being probabilities, can be neither negative nor greater than unity.

Earlier it was indicated that the artifact of difficulty factors arises in factor analysis when curvilinearities are present. Note that the derivation of the latent structure model embodies no restriction on curvilinear relations, either among the manifest attributes or between them and the latent dimensions.

The latent structure model has already shown promise in empirical research [cf. 14, 17]. Only its restriction to the analysis of dichotomous attributes prevents it from being a feasible alternative to the factor analysis of quantitative variables. The latent profile model to be taken up next is the generalization of latent structure analysis to the case of quantitative manifest variables [cf. 12].

Some Linear Recruitment Equations for Quantitative Manifest Variables

A reasonable question to ask is whether the natural linearity of recruitment equations could not provide a basis for analyzing interrelations among manifest variables that are quantitative rather than qualitative. This section will display such a system of recruitment equations.

Suppose there is available a set of s quantitative measures, such as test scores, on a sample of n people. On some basis let every sample member be assigned to one, and only one, of q subgroups. Then the size, sums of scores, and sums of score products for the entire sample are related to corresponding subgroup statistics in the following simple way.

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q , \\
\sum^{n} X_{ij} &= \sum^{n_1} X_{ij} + \sum^{n_2} X_{ij} + \cdots + \sum^{n_q} X_{ij} , \\
\sum^{n} X_{ij}X_{ik} &= \sum^{n_1} X_{ij}X_{ik} + \sum^{n_2} X_{ij}X_{ik} + \cdots + \sum^{n_q} X_{ij}X_{ik} , \\
\sum^{n} X_{ij}X_{ik}X_{il} &= \sum^{n_1} X_{ij}X_{ik}X_{il} \\
&\quad + \sum^{n_2} X_{ij}X_{ik}X_{il} + \cdots + \sum^{n_q} X_{ij}X_{ik}X_{il} , \quad \text{etc.}
\end{aligned}
\tag{6}
$$

All summations in (6) are over individuals. The summations on the left are over the entire sample, while those on the right are over the members of the various subgroups. X_{ij} is the score of individual i on test j, and it may be in raw, deviational, standard, or any other derived units. The same is true of X_{ik} , X_{il} , etc.

It is convenient to designate the various summations in (6) by the letter m with appropriate subscripts. Thus m_j and m_{jk} represent, respectively, the sum of scores on test j and the sum of products of scores on tests j and k, each sum being for the entire sample. m_{1j} and m_{1jk} , on the other hand, stand for the same things in subgroup 1 only, and so on. Equations (6) then become

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q , \\
m_j &= m_{1j} + m_{2j} + \cdots + m_{qj} , \\
m_{jk} &= m_{1jk} + m_{2jk} + \cdots + m_{qjk} , \\
m_{jkl} &= m_{1jkl} + m_{2jkl} + \cdots + m_{qjkl} , \quad \text{etc.}
\end{aligned}
\tag{7}
$$

The m's in (7) will be referred to as *product moments* of various orders. The product moments in the third line, involving two tests, will be said to be of second order, while those for three tests are of third order, and so on. For consistency, the m's involving only one test j will be called first-order product moments, and n, the sample size, could be called the zero-order product moment for the sample. The m's on the left in (7) will be called sample product moments, while those on the right are subgroup product moments.

It should be stressed that (6) and (7) in no way restrict either the empirical data or the method of classification into subgroups. It will be the burden of the next two sections to establish a basis for grouping that will convert these recruitment equations into a mathematical model.

The Fundamental Theorem of Latent Profile Analysis

Consider now a two-dimensional joint distribution in which the score of every person in the sample on test k is plotted against his score on test j.

Such a scatterplot is pictured in Figure 1. The unit of measurement for both tests is entirely arbitrary here. The ellipse in Figure 1 indicates only one of the possible shapes that the total configuration of points could have. The circle represents a subgroup t, of size n_t, *within* which the correlation between tests j and k is zero. The two lines labeled X_{tj} and X_{tk} in Figure 1 indicate the means of subgroup t on tests j and k. They intersect in what is called the *centroid* of the points comprising subgroup t. The point corresponding to individual i, a member of subgroup t, is shown with the coordinates (X_{ij}, X_{ik}), his scores on the two tests. The two distances, d_{ij} and d_{ik}, are the deviations of individual i from the j and k means of his subgroup. A property of such deviations is that their sum, over all members of the subgroup, is zero.

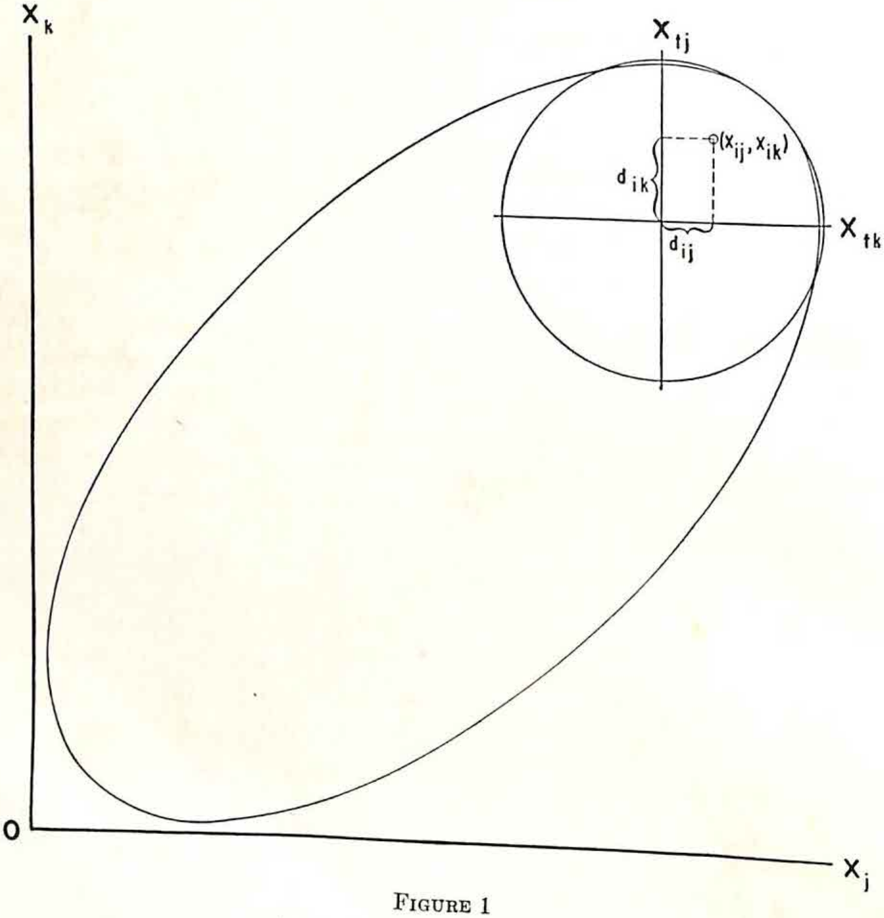


FIGURE 1
A Hypothetical Scatter Diagram

The j and k scores of individual i may be expressed in terms of his subgroup means and his deviations from those means, as follows:

$$X_{ij} = X_{tj} + d_{ij}\ , \tag{8}$$

$$X_{ik} = X_{tk} + d_{ik}\ . \tag{9}$$

Then m_{tj}, the sum of j scores for subgroup t, is given by

$$\begin{aligned} m_{tj} &= \sum^{n_t} X_{ij} \\ &= \sum^{n_t} (X_{tj} + d_{ij}) \\ &= \sum^{n_t} X_{tj} + \sum^{n_t} d_{ij} \\ &= n_t X_{tj}\ . \end{aligned} \tag{10}$$

The first term in the third line simplifies because X_{tj} is the same for every member of the subgroup. The last term in the third line vanishes because it is a sum of deviation scores. Thus the j scores for any subgroup t contribute to the sum of all j scores *as if* the members of the subgroup were concentrated at their mean for test j. This result (by no means a new one) will next be generalized, in an appropriate way, to the second-order product moment, m_{tjk}, for subgroup t.

By definition and with the help of (8) and (9), m_{tjk} becomes

$$\begin{aligned} m_{tjk} &= \sum^{n_t} X_{ij} X_{ik} \\ &= \sum^{n_t} (X_{tj} + d_{ij})(X_{tk} + d_{ik}) \\ &= \sum^{n_t} (X_{tj} X_{tk} + X_{tj}\, d_{ik} + X_{tk}\, d_{ij} + d_{ij}\, d_{ik}) \\ &= \sum^{n_t} X_{tj} X_{tk} + X_{tj} \sum^{n_t} d_{ik} + X_{tk} \sum^{n_t} d_{ij} + \sum^{n_t} d_{ij}\, d_{ik} \\ &= n_t X_{tj} X_{tk}\ . \end{aligned} \tag{11}$$

The first term in the fourth line of (11) simplifies because X_{tj} and X_{tk} are the same for all members of the subgroup. The second and third terms in that line vanish because they contain sums of deviations. The last term vanishes because it is the numerator of the formula for the correlation, within the subgroup, between tests j and k. The subgroup was earlier defined as having that correlation equal to zero. Thus the second-order product moment, m_{tjk}, is the same as it would be if all members of subgroup t were concentrated at their centroid. This holds for any subgroup within which tests j and k are uncorrelated.

A further distinction in terminology needs to be made here. The quantity

m_{tjk} is a sum of products of horizontal and vertical distances from the origin in Figure 1. It will therefore be called a *product moment about the origin.* Another kind of product moment in (11) is the last term in line 4, the sum of products of horizontal and vertical distances of a set of points from their own centroid. Such a quantity is appropriately called a *product moment about the centroid* of the set of points that is involved. The geometric interpretation of scores and deviations as distances makes it clear that the point of reference (origin, centroid, or some other point) of a product moment must always be specified in some way. Naturally this holds for all orders of product moments. Up to now, with four exceptions, all product moments have had the origin as their point of reference. The four exceptions are the last term in line 3 of (10) and the summations in the last three terms of line 4 in (11), which have the centroid of subgroup t as their point of reference.

The results of (10) and (11) can be generalized to higher order product moments by imposing additional restrictions on subgroup t. Not only must that subgroup be defined as having all pairs of tests j, k, and l uncorrelated within it, but, let it also have, for those three tests, a vanishing third-order product moment about its centroid. With these restrictions its third-order product moment about the origin, m_{tjkl} , becomes

(12)
$$m_{tjkl} = n_t X_{tj} X_{tk} X_{tl} \;.$$

This result is, in form and mode of development, a third-order analogue of the final step in the previous two equations. The fourth-order equivalent is obtained by analogous higher order restrictions, and so on.

Since in all of this discussion the origin could have been placed at any point 0, it now becomes possible to state the fundamental theorem of latent profile analysis.

> The g-order product moment, about any point O, of n_t points having zero product moments of order g and less about their centroid, is equal to the g-order product moment, about the point O, of n_t points placed at that centroid.

The Basic Equations of Latent Profile Analysis

The foregoing theorem provides a basis for grouping in the recruitment equations introduced previously. The close analogy with latent structure analysis will be obvious. Each subgroup or latent class should be homogeneous in whatever underlying dimensions are necessary to account for the observed interrelations. The homogeneity need not be complete, so long as deviations from the class averages are random, i.e., independent.

In the statistics of dichotomous attributes (as employed in coin-tossing experiments, for example), the notion of independence has usually applied to all orders of joint occurrence, and not just to pairs of events. This is the case in latent structure analysis, where within-class independence is *defined* as pertaining not only to all pairs of items but also to all triplets, all quadruplets, and so on. (It is easily shown by example that higher order independence among dichotomous attributes is *not* a mere logical consequence of uncorrelatedness between all pairs of such attributes.)

The concept of uncorrelatedness among quantitative measures, on the other hand, has more often been restricted, at least among psychometricians, to pairs of such measures. (A statistically oriented prepublication reviewer has pointed out that the two kinds of independence being discussed here are known to statisticians as *pair-wise* and *mutual* independence.) This is not an intrinsic or logical difference between qualitative and quantitative statistics. It is rather only a historical accident that the question of higher order independence among quantitative measures has less often arisen in psychometrics. That question arises here, for it turns out that the proper definition of such independence is crucial for this model.

In the previous section the within-class uncorrelatedness between pairs of tests was shown to be synonymous with vanishing second-order product moments about the centroid of the class. This is because such product moments are the numerators of the formulas for the corresponding correlations. Purely by analogy, higher order within-class independence may be equated with the vanishing of higher order product moments about the centroid of the class. (The failure of such product moments to vanish would allow, for example, a positive correlation between tests j and k among class members with high scores on test l, accompanied by a compensating negative correlation between the same two tests among class members having low scores on test l. This could happen in spite of zero correlations between all pairs of the three tests within the class as a whole. If correlational patterns can differ within subdivisions of a class, then the class is not homogeneous even from a commonsense point of view.) Therefore let the within-class independence of the present model be *defined* as applying to all orders of interrelations, and as expressing itself in the vanishing of product moments of all orders about the centroid of the class. Then the fundamental theorem applies with full force to the product moments of each class, so that the results of equations (10), (11), and (12) can be used to transform (7) into the basic equations of latent profile analysis:

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q\,, \\
m_j &= n_1 X_{1j} + n_2 X_{2j} + \cdots + n_q X_{qj}\,, \\
m_{jk} &= n_1 X_{1j} X_{1k} + n_2 X_{2j} X_{2k} + \cdots + n_q X_{qj} X_{qk}\,, \\
m_{jkl} &= n_1 X_{1j} X_{1k} X_{1l} + n_2 X_{2j} X_{2k} X_{2l} + \cdots + n_q X_{qj} X_{qk} X_{ql}\,, \quad \text{etc.}
\end{aligned}
\tag{13}
$$

Thus for s tests the 2^s manifest product moments (including n) are accounted for in terms of $(q + sq)$ latent parameters—the q latent class sizes and the

q class averages for each of the s tests. Each latent class is therefore characterized by its size and its profile of s test averages, its *latent profile.*

The latent profile equations and those of latent structure analysis turn out to be identical in form, as can be seen from a comparison of (5) and (13). This means that all algebraic latent structure solutions [1, 7, 11] are directly applicable to the latent profile equations. Hence the latent profile equations have a solution that can be obtained without the involvement of communality analogues (m_{jj}, m_{jjk}, m_{jjj}, etc.), and that is, in general, rotationally unique. (A conversation with Robert P. Abelson at Yale University has clarified the fact that, when deviational scores are used, the m_{jj} are the between-groups variances.) Nor do the latent profile equations restrict the occurrence of curvilinear relations among tests or between tests and underlying dimensions. Thus the dilemma of difficulty factors is avoided in this model.

Two Special Cases of Latent Profile Analysis

In the development of the latent profile equations the score units were entirely arbitrary. There are, however, two kinds of test scores that deserve special attention. Consider first the case where the manifest variables are, as in latent structure analysis, dichotomous attributes. Let the presence and absence of each such attribute be designated by scores of one and zero, respectively. In this case the manifest latent profile m's become identical with the manifest latent structure n's, and the class averages of latent profile analysis become the latent probabilities of latent structure analysis. The latent class sizes mean the same thing in both models. Thus the latent structure model is interpretable as the special case of latent profile analysis in which the manifest variables are dichotomous.

A second special result is obtained by using standard scores and by dividing the latent profile equations through by n, the number of people in the sample. The latent profile equations then assume their standard form:

$$\begin{aligned} 1 &= p_1 + p_2 + \cdots + p_q , \\ 0 &= p_1 Z_{1j} + p_2 Z_{2j} + \cdots + p_q Z_{qj} , \\ r_{jk} &= p_1 Z_{1j} Z_{1k} + p_2 Z_{2j} Z_{2k} + \cdots + p_q Z_{qj} Z_{qk} , \\ r_{jkl} &= p_1 Z_{1j} Z_{1k} Z_{1l} + p_2 Z_{2j} Z_{2k} Z_{2l} + \cdots + p_q Z_{qj} Z_{qk} Z_{ql} , \quad \text{etc.} \end{aligned} \tag{14}$$

The p's are the proportionate class sizes, and of course their sum is unity. The Z's are the average standard scores of classes on tests. Their weighted sum for any test (in the second line) vanishes because it is the mean of all standard scores on that test. The r_{jk}, being average products of pairs of standard scores, are the same correlations for which factor analysis attempts to account. The r_{jkl} are, analogously, average triple products, and so on.

The latent profile equations in standard form have the advantage of dealing with magnitudes that are independent of sample size and of arbitrary score units. It is in standard form that the equations will be applied to the examples in the next four sections.

Latent Profile Example I: *A Fictitious Two-Class Case*

The fictitious manifest data in Tables 1 and 2 will serve as a first latent profile example. Imagine the data as resulting from the administration of four alternate forms of the same test (such as arithmetic) to a sample of, let us say, a hundred people. Suppose, further, that all six intercorrelations turn out to be .50, so that every two-dimensional scatter diagram, with scores plotted in standard units, has the appearance of an ellipse twice as long as wide, centered on the origin, and tilted at an angle of 45 degrees. There are four three-dimensional scatter diagrams. Each is approximately egg-shaped, and, being symmetric about the origin, yields a third-order manifest product moment of zero.

Tables 1 and 2 display the necessary manifest data in a convenient way. The upper left entry in Table 1 is the first term in the first line of (14), the latent profile equations in standard form. The other entries in row and column 0 of Table 1 are the means of standard scores for each of the four tests—the left-hand term in the second line of (14). The remaining cells of Table 1 contain the test intercorrelations—the manifest data in the third line of (14).

Table 2 summarizes the manifest data of first, second, and third order. The upper left entry is the sum of the four means of standard scores. Each of the other entries in row or column 0 is the sum of the four correlations (including r_{jj}) involving the associated test. Every other cell in Table 2 contains the sum of the four third-order manifest product moments (including r_{jjk} and r_{jkk}) for the corresponding pair of tests.

For convenience of exposition in both fictitious examples in this paper, all elements with repeated subscripts (such as r_{jj}, r_{jjk}, and r_{jjj}) are treated as known. Their values are, in fact, easily inferred from the simple form of the manifest data, but this would not be true generally, even for all sets of fictitious data. In this first example, all r_{jj} are .50, all r_{jjk} are zero, and all r_{jjj} are zero.

Tables 1 and 2 have been labeled R and R_1 respectively. This is convenient notation for any such display of manifest data, and it will be used in all examples. In any latent profile solution, a distinction must be made between the given and the fitted R and R_1, the latter pair of tables indicating what the former should be in order to be completely accounted for by the solution. In both fictitious examples in this paper, the given and the fitted manifest data, the latter computed from the solution by means of (14), are identical and hence need not be compared. In the two empirical examples,

TABLE 1

R for a Fictitious Two-Class Case

Test No.	Test Number 0	1	2	3	4
0	1.00	.00	.00	.00	.00
1	.00	.50	.50	.50	.50
2	.00	.50	.50	.50	.50
3	.00	.50	.50	.50	.50
4	.00	.50	.50	.50	.50

TABLE 2

R_1 for a Fictitious Two-Class Case

Test No.	Test Number 0	1	2	3	4
0	.00	2.00	2.00	2.00	2.00
1	2.00	.00	.00	.00	.00
2	2.00	.00	.00	.00	.00
3	2.00	.00	.00	.00	.00
4	2.00	.00	.00	.00	.00

TABLE 3

Latent Profile Solution for a Fictitious Two-Class Case

	Test No.	Latent Class I	II
Class Means	1	-.71	.71
	2	-.71	.71
	3	-.71	.71
	4	-.71	.71
Class Sizes		.50	.50

TABLE 4

Given Correlations for Nine Reading Tests

Test No.	Test Number 0	1	2	3	4	5	6	7	8	9
0	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
1	.00	...	.72	.41	.28	.52	.71	.68	.51	.68
2	.00	.72	...	.34	.36	.53	.71	.68	.52	.68
3	.00	.41	.34	...	.16	.34	.43	.42	.28	.41
4	.00	.28	.36	.16	...	.30	.36	.35	.29	.36
5	.00	.52	.53	.34	.30	...	.64	.55	.45	.55
6	.00	.71	.71	.43	.36	.64	...	.76	.57	.76
7	.00	.68	.68	.42	.35	.55	.76	...	.59	.68
8	.00	.51	.52	.28	.29	.45	.57	.59	...	.58
9	.00	.68	.68	.41	.36	.55	.76	.68	.58	...

TABLE 5

Approximate Latent Profile Solution for Nine Reading Tests

	Test No.	Latent Class I	II
Class Means	1	-.80	.80
	2	-.81	.81
	3	-.47	.47
	4	-.41	.41
	5	-.68	.68
	6	-.90	.90
	7	-.85	.85
	8	-.66	.66
	9	-.84	.84
Class Sizes		.50	.50

however, the comparison will be made when possible in order to appraise the adequacy of the solution.

The latent profile solution for the present example, obtained by applying the algebra of the latent structure solution of Green [11], is shown in Table 3. Each of the two latent classes is defined in terms of its relative size and its latent profile—the complete set of average test scores for its members. Apparently Class I consists of those who are poor at arithmetic, while Class II contains the good arithmetic students.

A fruitful way to visualize this latent profile solution is in terms of the regressions of the tests on the latent continuum of arithmetic ability. Such a graph of mean test score, Y, against position along the latent continuum, X, is shown in Figure 2. Here the regressions of all four tests on the latent continuum are identical. In Figure 2 both Y and X are expressed in standard units, so that the slope of the regression line is also the correlation between Y and X. This is a correlation between test and "factor," and for the linear regressions of the present example, these correlations turn out to be exactly

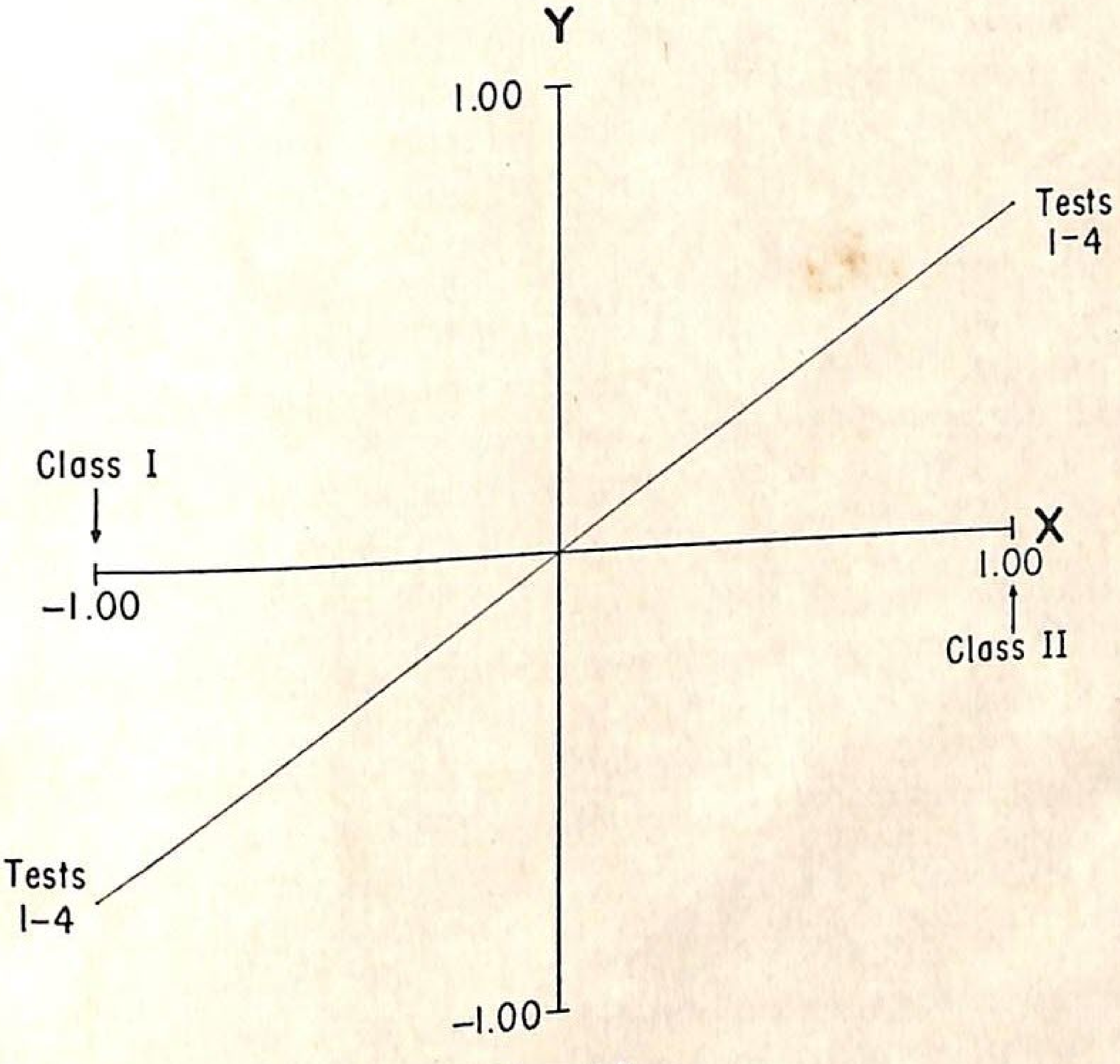


FIGURE 2

Regression of Tests on Latent Continuum for A Fictitious Two-Class Case the same ($\sqrt{2}/2$) as the factor loadings that would result from a factor analysis of the correlations in Table 1. This simple correspondence between the two models will vanish, however, as soon as any of the regressions become nonlinear.

Latent Profile Example II: *An Empirical Two-Class Case*

As a second latent profile example, consider the given R in Table 4. That table is a simple modification of a table of intercorrelations among nine reading tests previously reported by Davis [3]. The only modification was to border Davis' table with the 0 column and row. The intercorrelations were based on 421 cases.

A letter from Davis has indicated that the raw scores that would be needed for the computation of R_1 are no longer available. In the absence of higher order data which would provide a unique solution, it was necessary here to adapt some factoring and rotational procedures that were involved in an early approximate latent structure solution [5, 6, or 8] in order to obtain an approximate latent profile solution. For this purpose a factorization of the Davis data by Thurstone [19] was used. Thurstone's analysis indicated that one factor was sufficient to account for the data, all but three of the discrepancies between given and fitted correlations being less than .04, and the largest being .07. It is with exactly these same discrepancies that the present latent profile solution accounts for the intercorrelations.

An approximate latent profile solution for the Davis data is shown in Table 5. This solution is obtainable by resolving the rotational problem with any one of the following three assumptions:

(i) that the two latent classes are equal in size;

(ii) that the two latent profiles are identical except for reversed algebraic sign; or

(iii) that the given R_1, if available, would be like that of Example I in having nonzero entries only in its 0 column and row.

The solution in Table 5 generates, by means of (14), a fitted R_1 having the form indicated in assumption 3.

The regressions for this solution are pictured in Figure 3. Again both axes are in standard units, so that the slopes of the regressions are identical with the factor loadings reported by Thurstone. Although the present latent profile solution is not rotationally unique, it can fairly readily be shown to possess an important kind of invariance, namely, that *the slopes of the regressions, when both axes are in standard units, remain constant regardless of how the rotational indeterminacy is resolved.* This is but one illustration of the fact that the factor analysis and latent profile models are mutually complementary when the assumptions of both models are not violated.

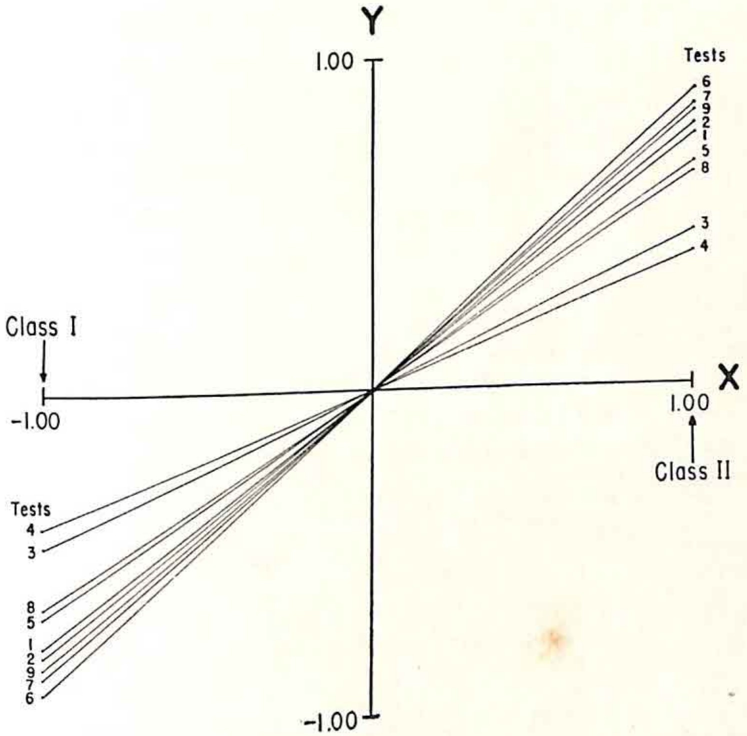


FIGURE 3
Regressions of Tests on Latent Continuum for Nine Reading Tests

Latent Profile Example III: *A Fictitious Three-Class Case*

The fictitious R and R_1 in Tables 6 and 7 will provide a first illustration of how latent profile analysis handles the problem of difficulty factors. Imagine tests 1 and 2 as being two easy vocabulary tests, tests 4 and 5 as two hard vocabulary tests, and 3 as a vocabulary test of intermediate difficulty. Again assume the data are based upon a hundred cases.

Before proceeding to the latent profile analysis of this fictitious data, it will be instructive to examine the results of a factor analysis of the correlations in Table 6. The simple structure factor analytic solution with correlated factors is given in Table 8. The entries in that table are the correlations between the five tests and the two factors, A and B. The correlation between the two factors, r_{AB}, is .33. If the usual rules for interpreting factors were followed unquestioningly here, the conclusion would be that the two factors are knowledge of easy words, A, and knowledge of hard words, B, and that the two abilities are relatively independent. This is absurd.

The unique, perfectly fitting latent profile solution for this example, again obtained from R and R_1 by the same algebra as is used in the latent

TABLE 6

R for a Fictitious Three-Class Case

Test No.	Test Number 0	1	2	3	4	5
0	1.00	.00	.00	.00	.00	.00
1	.00	.75	.75	.50	.25	.25
2	.00	.75	.75	.50	.25	.25
3	.00	.50	.50	.50	.50	.50
4	.00	.25	.25	.50	.75	.75
5	.00	.25	.25	.50	.75	.75

TABLE 7

R_1 for a Fictitious Three-Class Case

Test No.	Test Number 0	1	2	3	4	5
0	.00	2.50	2.50	2.50	2.50	2.50
1	2.50	-2.50	-2.50	-1.25	.00	.00
2	2.50	-2.50	-2.50	-1.25	.00	.00
3	2.50	-1.25	-1.25	.00	1.25	1.25
4	2.50	.00	.00	1.25	2.50	2.50
5	2.50	.00	.00	1.25	2.50	2.50

TABLE 8

Simple Structure Factor Analysis Solution for Correlations in Table 6

Test No.	Factors A	B
1	.82	.00
2	.82	.00
3	.41	.41
4	.00	.82
5	.00	.82

$r_{AB} = .33$

TABLE 9

Latent Profile Solution for a Fictitious Three-Class Case

	Test No.	Latent Class I	II	III
Class Means	1	-1.50	.50	.50
	2	-1.50	.50	.50
	3	-1.00	.00	1.00
	4	- .50	-.50	1.50
	5	- .50	-.50	1.50
Class Sizes		.25	.50	.25

structure solution of Green [11], is shown in Table 9. Figure 4 shows, in standard units, the regressions implied by Table 9 and by the assumption of equal spacing of the latent classes along the single latent continuum of vocabulary knowledge.

The contour of the various regressions in Figure 4 is exactly what would be expected on the basis of the relative difficulty of the tests. The easy tests (1 and 2) discriminate only at the lower end of the latent continuum. The hard tests (4 and 5) differentiate only at the upper end. Test 3, of medium difficulty, discriminates throughout the range.

Latent Profile Example IV: *An Empirical Three-Class Case*

The data in Table 10 will provide a final latent profile example. Aside from its 0 column and row, that table is merely a rounded version of a table of subtest intercorrelations reported by Ferguson ([4], p. 328) to illustrate the occurrence of difficulty factors. Ferguson took single items from a Moray-House verbal intelligence test and combined them into six subtests that were reasonably homogeneous in content but that increased in difficulty from subtest 1 to subtest 6. The correlations were based on a sample of 108 children, age 11.

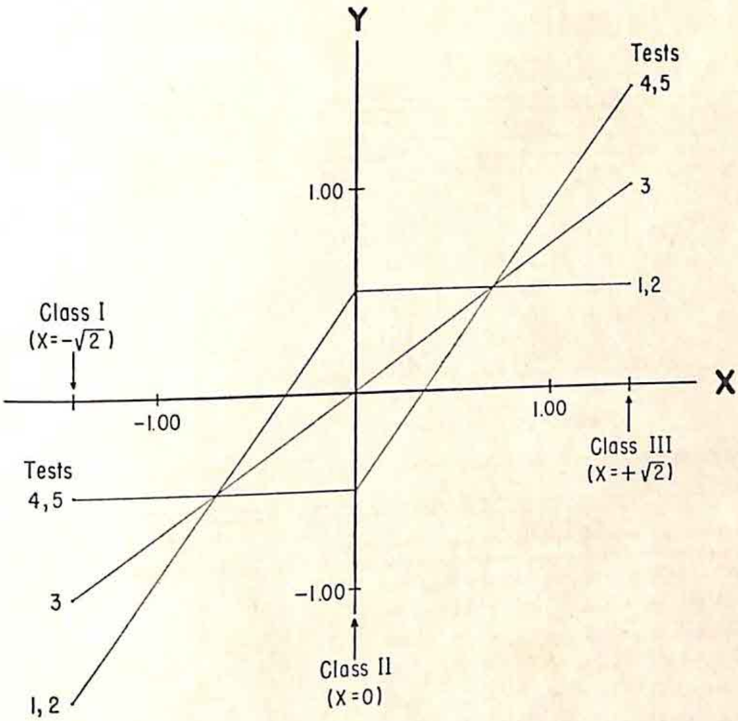


FIGURE 4
Regressions of Tests on Latent Continuum for A Fictitious Three-Class Case

The simple structure correlated factor solution for the Ferguson example is shown in Table 12. That solution was obtained by a rotation of the factorization reported by Ferguson ([4], p. 328). Here again it would be absurd to conclude, according to the usual rules, that the data must be thought of in terms of two relatively independent factors—high-level and low-level verbal intelligence.

A letter from Ferguson has indicated that the raw scores are no longer available for the computation of higher order manifest product moments. It therefore was necessary to approximate the latent profile solution by means of procedures similar to those used in the Davis example. For this purpose Ferguson's factorization of his correlations was used. His two factors accounted for the correlations with discrepancies not exceeding .03, and the present latent profile solution fits the correlations in eactly the same way.

The rotational indeterminacy here was resolved in two stages. The first step was to rotate Ferguson's factorization, with attention being given only to subtests 1 and 6, into maximum correspondence with the initial factorization for the first and last tests in Example III. The latter factorization was a part of the latent profile solution for that example. In Example III the unique solution was obtained by applying, to the initial factorization

TABLE 10

Given Correlations for Six Verbal Intelligence Sub-Tests

Test No.	0	1	2	3	4	5	6
0	1.00	.00	.00	.00	.00	.00	.00
1	.00	...	.86	.81	.81	.61	.37
2	.00	.86	...	.80	.82	.68	.47
3	.00	.81	.80	...	.87	.80	.67
4	.00	.81	.82	.87	...	.80	.68
5	.00	.61	.68	.80	.80	...	.78
6	.00	.37	.47	.67	.68	.78	...

TABLE 12

Simple Structure Factor Analysis Solution for Correlations in Table 10

Test No.	Factors A	Factors B
1	.86	.00
2	.75	.15
3	.55	.43
4	.57	.42
5	.30	.64
6	.00	.81

$r_{AB} = .43$

TABLE 11

Fitted R_1 for Six Verbal Intelligence Sub-Tests

Test No.	0	1	2	3	4	5	6
0	.00	4.31	4.49	4.80	4.84	4.47	3.77
1	4.31	-4.42	-3.93	-2.98	-3.07	-1.74	- .24
2	4.49	-3.93	-3.37	-2.30	-2.39	-1.02	.45
3	4.80	-2.98	-2.30	- .99	-1.08	.33	1.75
4	4.84	-3.07	-2.39	-1.08	-1.16	.26	1.70
5	4.47	-1.74	-1.02	.33	.26	1.52	2.71
6	3.77	- .24	.45	1.75	1.70	2.71	3.57

TABLE 13

Approximate Latent Profile Solution for Six Verbal Intelligence Sub-Tests

	Test No.	Latent Class I	Latent Class II	Latent Class III
Class Means	1	-1.64	.51	.62
	2	-1.56	.36	.84
	3	-1.38	.07	1.24
	4	-1.41	.09	1.23
	5	-1.05	-.20	1.44
	6	- .60	-.48	1.56
Class Sizes		.25	.50	.25

of R, a rotation completely specified by the higher order data in R_1. The second rotational stage in the present approximate solution was to imitate the Example III solution by using exactly the same rotation. The consequences of this rotational solution are the following.

(i) The relative class sizes are the same as in Example III.

(ii) The regressions of all six subtests on the latent continuum are ascending.

(iii) Assuming equal spacing of the classes along the latent continuum, the regressions of subtests 1 and 6 have curvatures that are approximately equal but opposite in direction.

(iv) The form of the fitted R_1 is similar to that of the given and fitted R_1 of Example III.

The resulting approximate latent profile solution is given in Table 13, and the regressions, again in standard units, and assuming equal spacing of the three classes along the latent continuum, are shown in Figure 5. The progression of curvatures from the easiest to the hardest subtest is just what it ought to be. This progression was found to be quite invariant over a wide range of alternative approximate solutions. Several such solutions were

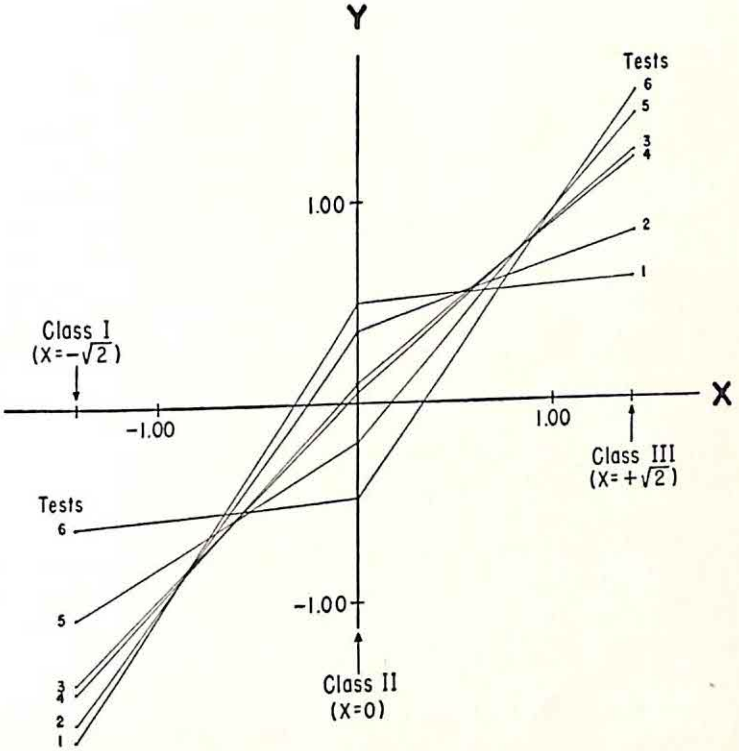


FIGURE 5
Regressions of Subtests on Latent Continuum for Six Verbal Intelligence Subtests

computed in order to study the nonuniqueness of this latent profile solution. Only one restriction applied to all of the alternative solutions that were tried, namely, that the resulting regressions be ascending for all subtests. Within this restriction large changes in class sizes and in class averages could be brought about, but never in such a way as to alter the ordering of curvatures among the regressions. It should be added that only with very strained assumptions about the spacing of the three classes along the latent continuum could the regressions of both subtests 1 and 6 be made to curve in the same direction.

Table 11 shows the fitted R_1 implied by the approximate solution in Table 13. A comparison of Tables 7 and 11 will reveal the similarities between the fitted higher order manifest data for the two three-class latent profile examples.

Discussion

There are two limitations to the foregoing latent profile examples that should be discussed explicitly. These are in addition to the indeterminacy produced by the absence of higher order manifest data in the case of the two empirical examples. The first is the lack of a scale of measurement for the latent continuum. For two classes this is unimportant, for it leaves only an arbitrariness as to the origin and the distances of the two classes from that origin. The problem of the relative distances between classes cannot arise when there is only one such distance. With three-class solutions, however, the problem of relative spacing is a critical one. Without some resolution of it the regressions of tests on the latent continuum could not be drawn. Nor could the shape of the distribution of positions along the latent continuum be ascertained. In both of the present three-class solutions, the problem was resolved by the *arbitrary* assumption of equal spacing of the classes along the latent continuum. The regressions were drawn on that basis, and on the same basis the latent distribution in each case became symmetric and approximately normal. Other assumptions about the underlying metric would have led to different regressions and to different latent distributions. A separate paper [9], stemming from some recent developments in latent structure analysis [16], deals further with this metric problem. It indicates one way in which, with the aid of manifest product moments of still higher order, a metric for the latent continuum can be made to emerge as an integral part of the latent profile solution.

A second limitation of all latent profile examples in the present paper is their unidimensionality. It will be recalled that nothing in the development of the latent profile equations restricts the number of underlying dimensions within which the latent classes lie. Of course the two-class examples here can be understood in terms of a single continuum, for that would be true of any two-class case. The present three-class examples, however, are unidimensional because of the special nature (homogeneous in content but graded in difficulty) of the tests involved in them. Many three-class examples would require two underlying dimensions for an adequate understanding of their psychological meaning. In general, a q-class solution could require as many as $(q - 1)$ underlying dimensions for its interpretation. Subsequent work [cf. 10] will deal with such multidimensional examples and with the problems of dimensionality and metric that arise in their interpretation.

The reader may have noticed that in all four latent profile examples no mention was made of a need for manifest data of order higher than the third. Even for the two empirical examples a unique solution would not have required the use of such higher order manifest data. These higher order data therefore constitute a means for testing the assumption of higher order within-class independence. This could be done by comparing the given higher order data with the corresponding fitted values as generated from the latent parameters by subsequent lines of equations (13) or (14). Alternatively it could be argued that, if the solution never requires data above third order, there is no need to postulate within-class independence beyond that order. From this parsimonious viewpoint, the latent profile equations need extend no higher than third order, and it would be inappropriate to think of using higher order data to test the fit of the model. However, it might then be more important to test the adequacy of the solution by relating it to variables not included in the original analysis. (An empirical example in [8] gives an illustration of how this could be carried out in latent structure analysis, where the very same argument over the use or non-use of higher order equations not needed in a particular solution can be made.)

The latent profile equations that have been derived and illustrated here are analogous to only one form of latent structure analysis—that known as the discrete class case. The analysis divides the sample into a small number of discrete classes possessing second- and third-order within-class independence, and stops there. Other varieties of latent structure analysis, one of which has already been referred to, have gone further in stipulating the algebraic form of the regressions (the so-called *trace lines* [cf. 15] of latent structure analysis) or of the set of class sizes, or of both. Always, however, the postulate of within-class independence is retained. Usually these further restraints require within-class independence of higher than third order, so that the corresponding higher orders of manifest data become directly involved in the solution. Most of these variants of latent structure analysis are readily translated into latent profile terms. In fact, the analogy between the two models is so close that almost whatever progress is made in the various solutions for one model is convertible into a corresponding advance for the other.

Conclusion

After outlining the derivation of the factor analysis and latent structure models, this paper has shown how the latter can be generalized for analyzing the interrelations among quantitative measures in a way that avoids some of the troublesome problems of factor analysis. The resulting latent profile model is applied to some simple fictitious and empirical data to illustrate its use. Because such applications may seem to show some promise, it is perhaps appropriate and not premature to conclude this paper merely by broadening the reference for the admonition ([18], p. 70) with which it began.

> It would be unfortunate if some initial success with the analytical methods . . . described here should lead us to commit ourselves to them with such force of habit as to disregard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

REFERENCES

[1] Anderson, T. W. On estimation of parameters in latent structure analysis. *Psychometrika*, 1954, **19**, 1-10.

[2] Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1-19.

[3] Davis, F. B. Fundamental factors of comprehension in reading. *Psychometrika*, 1944, **9**, 185-197.

[4] Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, **6**, 323-329.

[5] Gibson, W. A. Applications of the mathematics of multiple-factor analysis to problems of latent structure analysis. In P. F. Lazarsfeld et al., The use of mathematical models in the measurement of attitudes. Res. Memo. No. 455, Rand Corp., 1951.

[6] Gibson, W. A. Applications of the mathematics of multiple-factor analysis to problems of latent structure analysis. Unpublished doctoral dissertation, Univ. Chicago, 1951.

[7] Gibson, W. A. An extension of Anderson's solution for the latent structure equations. *Psychometrika*, 1955, **20**, 69-73.

[8] Gibson, W. A. Multiple factors and latent structure. Unpublished manuscript.

[9] Gibson, W. A. Nonlinear factor analysis, single factor case. *Amer. Psychologist*, 1955, **10**, 438. (Abstract)

[10] Gibson, W. A. Nonlinear factors in two dimensions. *Amer. Psychologist*, 1956, **11**, 415 (Abstract)

[11] Green, B. F. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 1951, **16**, 151-166.

[12] Green, B. F. Latent structure analysis and its relation to factor analysis. *J. Amer. statist. Ass.*, 1952, **47**, 71-76.

[13] Guilford, J. P. The difficulty of a test and its factor composition. *Psychometrika*, 1941, **6**, 67-77.

[14] Lazarsfeld, P. F. The interpretation and computation of some latent structures. In S. A. Stouffer et al., *Measurement and prediction*. Princeton: Princeton Univ. Press, 1950. Ch. 11.

[15] Lazarsfeld, P. F. The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer et al., *Measurement and prediction*. Princeton: Princeton Univ. Press, 1950. Ch. 10.

[16] Lazarsfeld, P. F. Some new results and problems in latent structure analysis. Unpublished manuscript.

[17] Lazarsfeld, P. F., et al. The use of mathematical models in the measurement of attitudes. Res. Memo. No. 455, Rand Corp., 1951.

[18] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

[19] Thurstone, L. L. Note on a reanalysis of Davis' reading tests. *Psychometrika*, 1946, **11**, 185-188.

[20] Wherry, R. J. and Gaylord, R. H. Factor pattern of test items and tests as a function of the correlation coefficient: content, difficulty, and constant error factors. *Psychometrika*, 1944, **9**, 237-244.

Manuscript received 2/7/58

Revised manuscript received 9/19/58

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

STRATEGIES AND LEARNING MODELS

STEVEN J. BRYANT AND JOHN G. MARICA

FRESNO STATE COLLEGE

A class of strategies is defined, each member of which possesses a certain plausibility. If a subject follows any strategy in this class in a two-choice learning experiment of the type dealt with by the Estes model, the subject's long-run behavior will be the same as that predicted by the Estes model.

Consider a partial reinforcement experiment in which there are two alternatives of behavior A_1 and A_2. If A_1 is chosen on a particular trial, reward occurs with probability π_1; if A_2 is chosen, reward occurs with probability π_2. Let $p_1(t)$ be the probability the subject chooses A_1 on trial t.

In special cases of the Estes stimulus-sampling model Estes ([1], ch. 9, p. 134) has found that if $\bar{p}_1(t)$ is the expected mean probability of choosing A_1 then

$$(1) \quad \bar{p}_1(n+1) = \theta(1-\pi_2) + (1 - 2\theta + \theta\pi_1 + \theta\pi_2)\bar{p}_1(n), \quad (\theta \text{ a constant})$$

and, letting $\bar{p}_1 = \lim_{t\to\infty} \bar{p}_1(t)$ then,

$$(2) \qquad \bar{p}_1 = \frac{1-\pi_2}{2-\pi_1-\pi_2}.$$

Since the subject can "do better" (in terms of maximizing his expectation) by always choosing the alternative having greater reward probability, questions arise as to the rationality of this type of behavior (Flood [1], ch. 18).

Simon [2] has shown that the Estes result can be derived from the assumption that the subject is behaving rationally in a certain game-theoretic sense and is attempting to minimize his "regret."

In this note,

$$\lim_{t\to\infty} p_1(t) = \frac{1-\pi_2}{2-\pi_1-\pi_2}$$

is derived from the assumption that the subject is adhering to one of a certain class of strategies $\{S_\gamma\}$. A member of this class may be described by having the subject decide in advance on a policy, i.e.:

(i) when a choice is followed by reward, the same choice will be made on the next trial *all the time* (with probability 1);

(ii) when a choice is not followed by reward, the same choice will be made on the next trial with some fixed probability $\gamma \neq 1$.

The above axioms may be interpreted in several ways, e.g., the subject equips himself with a mechanism such as a barrel of balls, γ of which are marked "repeat," and he samples the barrel with replacement each time he is not rewarded, to decide what to do on the next trial. Or the subject has some built-in mechanism which affects his behavior in the same way. At any rate, each real number γ, $0 \leq \gamma < 1$, determines a strategy, call it S_γ.

The equation

$$p_1(n) = \pi_1 p_1(n-1) + (1-\pi_1)p_1(n-1)\gamma + (1-\pi_2)[1-p_1(n-1)](1-\gamma) \tag{3}$$

may be regarded as a formal statement of axioms (i) and (ii), for if the subject is following strategy S_γ he will choose A_1 on trial n:

(a) if he chose A_1 on the trial $n-1$ and was rewarded, (which occurs with probability $\pi_1 p_1\,(n-1)$);

(b) if he chose A_1 on the trial $n-1$ and was not rewarded, and the mechanism tells him to choose A_1 on the next trial, (which occurs with probability $(1-\pi_1)p_1(n-1)\gamma$);

(c) if he chose A_2 on the trial $n-1$, was not rewarded, and the mechanism tells him to choose A_1 on the next trial, (which occurs with probability $(1-\pi_2)\,[1-p_1(n-1)]\,(1-\gamma)$).

The above difference equation is of the same type as the one obtained from the stimulus-sampling model and

$$\lim_{t\to\infty} p_1(t) = \frac{1-\pi_2}{2-\pi_1-\pi_2}. \tag{4}$$

This time, the limit is independent of γ and so each member S_γ of the class $\{S_\gamma\}$ yields the same limiting behavior in agreement with the Estes model as well as that of Bush, Mosteller, and Thompson ([1], ch. 8).

Except for the bars over $p_1(n)$ and $p_1(n-1)$, (1) is the same as (3) if and only if $\theta = 1 - \gamma$. From this, one may conclude that, if a subject follows one of the strategies S_γ defined by (1) and (2), his trial-by-trial behavior will be the same as the mean behavior predicted by the stimulus-sampling model if and only if γ, the probability of repeating after nonreward, is equal to $1-\theta$; to stretch a point, the constant value attributed to θ may be paraphrased "hope springs eternal, with probability $1-\theta$."

A Generalization

Each member of the class of strategies just indicated determines the subject's choice on the basis of what happened on the preceding trial. There are also strategies in which the subject takes into account what has happened on the preceding k trials, which also lead to the same asymptote.

To construct such strategies, let k be a fixed positive integer. Also let $w_1, \cdots, w_k$ be k nonnegative real numbers such that $\sum_{i=1}^{k} w_i = 1$. Let $\gamma_1, \cdots, \gamma_k$ be k real numbers such that $0 \leq \gamma_i < 1$. Let

$$\delta_i = \pi_1 p_1(n - i) + (1 - \pi_1)p_1(n - i)\gamma_i + (1 - \pi_2)[1 - p_1(n - i)](1 - \gamma_i).$$

Then let

$$p_1(n) = \sum_{i=1}^{k} w_i\, \delta_i\,.$$

This not only defines $p_1(n)$, the probability of choosing A_1 on the trial n, but yields a strategy, i.e., the subject looks at the last k trials, each of which has a "weight." Thus, if the subject chose A_1 on the trial $n - i$ and was rewarded, that trial contributes weight w_i. If the subject chose A_1 and was not rewarded on the trial $n - i$, then that trial contributes weight $w_i\gamma_i$. If the subject chose A_2 and was not rewarded on the trial $n - i$, that trial contributes weight $w_i(1 - \gamma_i)$. Then the probability of choosing A_1 on the trial n, given a particular sequence of k preceding outcomes, is the sum of the weights contributed. For example, if $\gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$, then each trial contributes weight w_i if the subject chose A_1 and was rewarded, or chose A_2 and was not rewarded; otherwise, the trial contributes 0 weight to the probability of choosing A_1 on the trial n.

Now, it is easy to show that

$$\lim_{n\to\infty} p_1(n) = \frac{1 - \pi_2}{2 - \pi_1 - \pi_2}.$$

Proof. $\lim_{n\to\infty} p_1(n)$ exists. Let $\lim_{n\to\infty} p_1(n) = p_1$; then p_1 is the solution of the equation

$$p_1 = \pi_1 p_1 \sum_{i=1}^{k} w_i + (1 - \pi_1)p_1 \sum_{i=1}^{k} w_i\gamma_i + (1 - \pi_2)(1 - p_2) \sum_{i=1}^{k} w_i(1 - \gamma_i).$$

Let

$$\gamma = \sum_{i=1}^{k} w_i\gamma_i\,;$$

then

$$p_1 = \pi_1 p_1 + (1 - \pi_1)p_1\gamma + (1 - \pi_2)(1 - p_1)(1 - \gamma)$$

and $\gamma \neq 1$, and hence

$$p_1 = \frac{1 - \pi_2}{2 - \pi_1 - \pi_2}.$$

This shows that, *if the subject follows a strategy S_γ in the class $\{S_\gamma\}$*, his asymptote will be the same regardless of how good his "memory" is. In short, he will do no better in the long run by considering, say; the last 100 trials each time, than if he only considers the last trial; and furthermore, his long-run behavior is unaffected by the choice of weights which he gives to each trial.

REFERENCES

[1] Thrall, R. M., Coombs, C. H., and Davis, R. L. (Eds.) *Decision processes*. New York: Wiley, 1954.

[2] Simon, H. A. A comparison of game theory and learning theory. *Psychometrika*, 1956, 21, 267-272.

Manuscript received 10/6/58

Revised manuscript received 1/12/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON TRYON'S MEASURE OF RELIABILITY

ALLEN L. EDWARDS

THE UNIVERSITY OF WASHINGTON

Two alternative formulas, based upon the analysis of variance, are given for Tryon's general form for the reliability coefficient.

A general form for the reliability coefficient r_{tt} of an unstratified composite X_t, consisting of the sum of n observations obtained for each of N individuals, has been published by Tryon [2]. For example, for an $N \times n$ matrix of observations, as shown in Table 1, with N individuals or rows and n columns of observations, the general form is given by

$$(1) \qquad r_{tt} = \frac{n^2\bar{c}_{ij}}{V_t},$$

where $\bar{c}_{ij}$ is the average of the covariances of the entries in the n columns, and V_t is the variance of the N row sums.

Tryon shows that r_{tt} may be calculated from any one of a number of algebraically identical formulas, but does not show the relationship between (1) and two useful alternatives based upon the analysis of variance. For example, (1) can also be written, in a form given by Hoyt [1], as

$$(2) \qquad r_{tt} = 1 - \frac{\text{MS}_{rc}}{\text{MS}_r},$$

where MS_{rc} is the row × column interaction mean square and MS_r is the mean square between rows, or as

$$(3) \qquad r_{tt} = \frac{n}{n-1}\left(1 - \frac{\text{MS}_{wc}}{\text{MS}_r}\right),$$

where MS_{wc} is the mean square within columns.

For the data of Table 1 Tryon reports r_{tt}, based upon (1), as .880. In Table 2 appears the analysis of variance for the same example. Substituting in (2) with the appropriate values from Table 2 yields

$$r_{tt} = 1 - \frac{3.728}{31.122} = .880.$$

For the same example, the sum of squares within columns is 280.1 +

TABLE 1*

Illustrative Score Matrix

Subjects	Test Samples					
	X_1	X_2	X_3	X_4	X_5	X_t
1	6	2	1	0	0	9
2	8	6	5	2	4	25
3	10	12	7	7	7	43
4	5	11	11	9	8	44
5	6	3	0	0	1	10
6	11	7	9	6	1	34
7	7	7	2	5	5	26
8	4	7	4	4	1	20
9	6	3	3	2	4	18
10	6	5	1	3	1	16
ΣX_i	69	63	43	38	32	245

*From Tryon [2]

TABLE 2

Analysis of Variance of the Scores Given in Table 1

Source of variation	Sum of squares	df	Mean square
Between rows (subjects)	280.1	9	31.122
Between columns (scores)	104.2	4	26.050
Row x column interaction	134.2	36	3.728
Total	518.5	49	

134.2 = 414.3, with $n(N-1)$ degrees of freedom. MS_{wc} is, therefore, equal to 414.3/45 = 9.207. Substituting with this value in (3),

$$r_{tt} = \frac{5}{5-1}\left(1 - \frac{9.207}{31.122}\right) = .880.$$

The two analysis of variance formulas will be shown to be algebraically equivalent to (1). Let the sum of squares within columns be $\sum x_{wc}^2$, the sum of squares between rows be $\sum x_r^2$, and the row × column sum of squares be $\sum x_{rc}^2$, with degrees of freedom of $n(N-1)$, $(N-1)$, and $(n-1)(N-1)$, respectively. Division of each sum of squares by its degrees of freedom gives the corresponding mean squares MS_{wc}, MS_r, and MS_{rc}.

The mean of the variances within columns is

$$\bar{V}_i = (s_1^2 + s_2^2 + \cdots + s_n^2)/n = \frac{\sum x_{wc}^2}{n(N-1)} = \text{MS}_{wc}. \tag{4}$$

Tryon's V_t is the variance of the row *sums* and, in his notation, is

$$V_t = n\bar{V}_i + n(n-1)\bar{c}_{ij}. \tag{5}$$

The variance of the row *means* will then be

$$s_{\bar{x}}^2 = \frac{1}{n^2}\,[n\bar{V}_i + n(n-1)\bar{c}_{ij}]. \tag{6}$$

Also, $ns_{\bar{x}}^2 = \text{MS}_r$ and, therefore,

$$n^2 s_{\bar{x}}^2 = n\,\text{MS}_r = V_t. \tag{7}$$

Solving for $\bar{c}_{ij}$ in (5) and substituting with this value in (1), gives Tryon's variance form or

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum V_i}{V_t}\right). \tag{8}$$

To derive formula (3), substitute from (4) and (7) in (8) to obtain

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\text{MS}_{wc}}{\text{MS}_r}\right). \tag{9}$$

Formula (2) may be derived by noting that $\sum x_{wc}^2 = \sum x_{rc}^2 + \sum x_r^2$. Then, substituting in (9)

$$\begin{aligned} r_{tt} &= \frac{n}{n-1} - \frac{(\sum x_{rc}^2 + \sum x_r^2)/(N-1)}{(n-1)\,\text{MS}_r} \\ &= 1 - \frac{\text{MS}_{rc}}{\text{MS}_r}. \end{aligned}$$

REFERENCES

[1] Hoyt, C. Test reliability obtained by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.

[2] Tryon, R. C. Reliability and behavior domain validity: reformulation and historical critique. *Psychol. Bull.*, 1957, **54**, 229-249.

Manuscript received 8/16/58

Revised manuscript received 2/9/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

AN IMPROVED PROCEDURE FOR THE WHERRY-WINER METHOD FOR FACTORING LARGE NUMBERS OF ITEMS

LEROY WOLINS

IOWA STATE COLLEGE

A technique is presented that differs from the previous one in that the use of variance terms is eliminated from the computations; thus some formulas are simplified. A rationale for the improved method is presented.

In a previous article [1] the following formula is recommended for successive use to secure estimates of factor loadings:

$$r_{iK} = \frac{\left(r_{iK'}\dfrac{\sigma_{K'}}{\bar{\sigma}_i} - 1\right) + h_{ii}^2}{\sqrt{\left(\dfrac{\sigma_{K'}^2}{\bar{\sigma}_i^2} - n_{K'}\right) + \sum h_{ii}^2}}. \tag{1}$$

In (1) r_{iK} is the factor loading of an item. The difference between $r_{iK'}$ and r_{iK} is that the latter is based on computations involving communalities whereas the former is based on the same correlation matrix but with unity entered in each diagonal cell. The standard deviation of a total score based on all of the items included in a cluster is $\sigma_{K'}$, and the average standard deviation of the items included in cluster K is $\bar{\sigma}_i$. The number of items included in cluster K is $n_{K'}$ and, of course, represents the number of ones in the diagonal of the matrix of item intercorrelations.

By making substitutions in (1) suggested by the relationships

$$\frac{\sigma_{K'}}{\bar{\sigma}_i} = \sqrt{\sum \sum r_{ij}}, \tag{2}$$

and

$$\sum r_{iK'} = \sum \left[\frac{\sum r_{ij}}{\sqrt{\sum \sum r_{ij}}} \right] = \sqrt{\sum \sum r_{ij}}, \tag{3}$$

(1) becomes

$$r_{iK} = \frac{(r_{iK'} \sum r_{iK'} - 1) + h_{ii}^2}{\sqrt{(\sum r_{iK'})^2 - n_{K'} + \sum h_{ii}^2}}. \tag{4}$$

As in the previous method the correlations of items not in the cluster are converted to factor loadings by multiplying them by $C_{K'}$ and

$$r_{KL} = C_{K'} C_{L'} r_{K'L'} , \tag{5}$$

but by the present method

$$C_{K'} = \frac{\sum r_{iK'}}{\sum r_{iK}} . \tag{6}$$

The rationale for this modification is discussed in connection with the following identity:

$$r_{aK'} = \frac{[r_{\sigma_a r_{aj}}][\sigma_{\sigma_i}][\sigma_{r_{aj}}] + \bar{\sigma}_i \bar{r}_{aj}}{\sqrt{[r_{(\sigma_i \sigma_j) r_{ij}}][\sigma_{\sigma_i \sigma_j}][\sigma_{r_{ij}}] + \bar{\sigma}_i^2 \bar{r}_{ij}}} , \tag{7}$$

where

$r_{aK'}$ = the correlation of a particular item, a, with the cluster that contains it,
$r_{\sigma_a r_{aj}}$ = the correlation of a particular column or row of item intercorrelations with the column or row of item standard deviations,
σ_{σ_i} = the standard deviations of the item standard deviations,
$\sigma_{r_{aj}}$ = the standard deviation of a particular row or column of item intercorrelations,
$\bar{\sigma}_i$ = the mean standard deviation of the items,
$\bar{r}_{aj}$ = the mean of a particular column or row of item intercorrelations,
$r_{(\sigma_i \sigma_j) r_{ij}}$ = the correlation of the product of all combinations and permutations of two item standard deviations with their respective item intercorrelations,
$\sigma_{\sigma_i \sigma_j}$ = the standard deviation of the product of all combinations and permutations of two item standard deviations,
$\sigma_{r_{ij}}$ = the standard deviation of all item intercorrelations,
$\bar{r}_{ij}$ = the mean of all item intercorrelations.

The pertinent aspect of (7) is that it illustrates how it might be reduced to the simplified form suggested in the present paper. Such simplification is realized when any of the three bracketed terms in both the numerator and denominator is zero or is near zero.

The first terms to consider are $\sigma_{r_{aj}}$ and $\sigma_{r_{ij}}$. With respect to manipulating the data so as to make these terms zero, nothing can be done. It is expected that the correlations within a correlation matrix or within a row or column of such a matrix will vary, but these standard deviations will certainly be less than one, and if the clustering was done well, these standard deviations will be quite small.

The σ_{σ_i} and $\sigma_{\sigma_i \sigma_j}$ terms can be manipulated. It is conceivable with the use of electronic computors that the distribution of item responses to each item could be economically transformed to standard scores prior to computing the sums to obtain cluster scores. Later comments will point out that this is probably not justified.

The $r_{\sigma_a r_{ai}}$ and $r_{(\sigma_i \sigma_j) r_{ij}}$ terms will be zero on the average, theoretically, if tetrachorics are used and, theoretically, will be positive on the average if phi coefficients are used, varying in size depending on σ_{σ_i} and $\sigma_{\sigma_i \sigma_j}$. However, even if chance were the only thing contributing to the variability of $r_{\sigma_a r_{ai}}$ and $r_{(\sigma_i \sigma_j) r_{ij}}$, the user of the present technique would not wish to depend on any theoretical average value since the standard error of these coefficients would be quite large with the small number of observations which would occur in those problems requiring iteration.

In the Wherry-Winer paper the assumption made was that σ_{σ_i} is zero. It was also assumed, tacitly, that the bracketed terms in the denominator of (7) are the same for tetrachorics as for phis. Since both $r_{\sigma r}$ and σ_r will change as a result of using tetrachorics, this latter tacit assumption is not justified; however, since all three terms in both the denominator and the numerator will be less than one and probably closer to zero than to one, the product of the three bracketed terms will probably be very small. This appears to be the reason Wherry and Winer found close correspondence of their method to results obtained through actually extracting a centroid. This becomes evident when one considers that if items included in a cluster vary in difficulty as much as from .2 to .8, σ_{σ_i} will be in the order of .02 and $\sigma_{\sigma_i \sigma_j}$ will be even smaller.

Thus, it is concluded on a rational basis that the present computational procedure is an improvement of the former one, certainly with respect to computational ease and certainly when tetrachorics are used. With respect to phis, it is difficult to determine if the present method is better or worse than the former one. One can say with certainty that there will not be much difference between them on the average since the product of the bracketed terms will be small in most cases.

A third method available, discussed in [2], is clearly superior to both methods discussed above, in that the results of this third method are identical to those obtained from extracting a centroid from a cluster. An objection to the general use of this exact method is that it requires more work. The exact method should be used on those occasions when iteration leads to divergence and the investigator wishes to salvage the cluster. There is no guarantee that the exact method will salvage the cluster, of course, since divergence can be caused by a negative r_{ij} as well as by large values of the bracketed terms.

Another use of the exact method is when the items are not dichotomous. If items are responded to on a 1 to 5 scale, say, it may be easier to use the exact method than to dichotomize.

The computational formula for the exact method is

$$\hat{r}_{jK} = \frac{\sigma_j(r_{jK'} \sum \sigma_i r_{iK'} - \sigma_j) + \hat{h}^2_{jj}}{\sqrt{\sum [\sigma_j(r_{jK'} \sum \sigma_i r_{iK'} - \sigma_j) + \hat{h}^2_{jj}]}}, \tag{8}$$

where

$$r_{jK} = \frac{\hat{r}_{jK}}{\sigma_j}. \tag{9}$$

REFERENCES

[1] Wherry, R. J. and Winer, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, **18**, 161-179.

[2] Winer, B. J. Iterative factor analysis: its psychological and mathematical bases. Unpublished doctoral dissertation, Ohio State Univ., 1952.

Manuscript received 3/27/58

Revised manuscript received 3/11/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

GENERATING VARIABLES WITH ARBITRARY PROPERTIES

PAUL J. HOFFMAN

UNIVERSITY OF OREGON

There are occasions in psychological research where it is desirable to have available sets of variables with arbitrary intercorrelations. A quite simple procedure is described for generating pairs of such variables.

There are a number of instances in research and teaching where it is desired to produce fictitious data that exhibit particular characteristics. Linear transformation of scores on an existing variable can easily produce a variable with any desired mean and standard deviation. In test construction, where item information is available, the items can be so selected as to yield a test of given difficulty and reliability [1]. It also has been shown [2] that a test can be constructed to yield a given weight in relation to a second test or a composite.

This note describes a procedure whereby pairs of variables may be constructed from a table of random numbers so that their correlation will be of any given predetermined magnitude. The operations can be easily carried out in a few minutes on a desk calculator.

Let X and Z be random variables with the restriction $r_{xz} = 0$. We wish to determine a distribution Y such that r_{xy} is of some arbitrary size. Let $Y_i = X_i + bZ_i$. In deviation score notation,

$$r_{xy} = \sum_{i=1}^{N} \frac{(x_i)(x_i + bz_i)}{N\sigma_x\sigma_{(x+bz)}} = \frac{\sigma_x^2 + br_{xz}\sigma_x\sigma_z}{\sigma_x\sigma_{(x+bz)}}$$

which, with $r_{xz} = 0$ becomes

$$r_{xy} = \frac{\sigma_x}{\sigma_{(x+bz)}}. \tag{1}$$

But

$$\sigma_{x+bz}^2 = \sigma_x^2 + b^2\sigma_z^2 . \tag{2}$$

Therefore, by substitution,

$$r_{xy}^2 = \frac{\sigma_x^2}{\sigma_x^2 + b^2\sigma_z^2},$$

or, solving for b,

$$b^2 = \frac{\sigma_x^2(1 - r_{xy}^2)}{\sigma_z^2 r_{xy}^2}, \qquad b = \frac{\sigma_x}{\sigma_z}\sqrt{\frac{1 - r_{xy}^2}{r_{xy}^2}};$$

since $\sqrt{1-r_{xy}^2} = k_{xy} =$ the coefficient of alienation,

$$b = \frac{k_{xy}\sigma_x}{r_{xy}\sigma_z}, \tag{3}$$

from which the Y_i can be readily computed.

The Y distribution which results will have a mean,

$$\bar{Y} = \bar{X} + b\bar{Z}$$

TABLE 1

Two Normal Variables, X and Z. $\overline{X} = \overline{Z} = 5.00$, $\sigma_X = \sigma_Z = 1.00$, and $r_{XZ} = .00$

X	Z	X	Z	X	Z	X	Z
5.966	4.249	5.424	5.404	5.188	5.303	5.044	3.711
5.613	5.741	6.236	4.793	8.044	4.852	4.357	5.166
5.339	4.452	3.979	6.024	4.940	6.163	4.989	4.487
2.539	3.733	4.388	5.726	4.709	5.449	6.942	4.794
4.414	5.851	5.160	6.472	3.712	5.460	4.351	4.921
5.727	3.781	5.499	5.112	3.598	3.677	2.854	5.176
5.334	6.403	6.270	5.068	3.595	5.494	5.402	6.906
4.615	7.093	6.166	5.053	5.821	5.635	4.425	5.104
5.953	4.690	5.078	6.324	3.529	4.556	4.325	5.603
6.527	6.158	5.533	2.699	4.269	5.120	3.818	6.525
5.061	3.571	4.174	5.414	5.691	3.583	5.808	3.862
5.280	4.093	2.927	6.493	4.966	4.904	6.174	3.949
5.041	5.088	5.003	4.229	4.993	4.304	5.722	6.794
2.973	4.690	5.792	5.720	4.893	6.666	3.816	5.129
5.169	4.966	6.186	5.442	3.880	5.436	4.129	5.458
3.714	4.158	4.364	6.345	5.511	5.737	5.097	4.593
5.929	3.976	3.789	5.193	4.853	5.269	3.866	3.202
6.138	6.024	3.671	4.795	4.473	3.288	5.196	6.876
5.832	5.868	5.704	3.971	6.256	5.371	4.191	5.037
6.660	4.126	4.365	5.208	4.891	3.955	5.503	3.845
5.510	3.924	4.768	6.919	3.756	4.723	6.346	5.028
5.030	3.810	3.949	3.760	5.069	3.927	4.278	5.138
5.983	3.050	5.738	5.736	3.728	4.463	4.669	4.980
4.849	5.119	7.519	5.064	5.575	4.773	6.041	5.607
5.586	6.887	4.657	4.397	4.156	2.418	5.440	4.722

and a standard deviation,

$$\sigma_y = \sqrt{\sigma_x^2 + b^2\sigma_z^2}\ .$$

If it is desired that the Y distribution have some arbitrary mean, $\bar{Y}_A$, and standard deviation, σ_{yA}, as well as an arbitrary correlation with X, the individual scores may be computed from the formula

$$Y_i = \frac{\sigma_{yA}}{\sigma_y}(X_i + bZ_i) + C$$

where σ_{yA} = the desired standard deviation,
σ_y = the obtained standard deviation, $\sqrt{\sigma_x^2 + b^2\sigma_z^2}$

$$C = \bar{Y} - \frac{\sigma_{yA}}{\sigma_y}(\bar{X} + b\bar{Z})$$

and b is defined as in (3) above.

In the special case, where

$$\bar{X} = \bar{Z}, \qquad \sigma_x = \sigma_z = 1.00, \quad \text{and} \quad r_{xz} = 0,$$

we find by substitution that

$$b = k_{xy}/r_{xy}\ , \qquad \bar{Y} = \bar{X}(1 + [k_{xy}/r_{xy}]), \qquad \sigma_y = 1/r_{xy}\ .$$

Table 1 contains, for $N = 100$, a sample pair of normal variables (X and Z) such that $\bar{X} = \bar{Z} = 5.00$, $\sigma_x = \sigma_z = 1.00$, and $r_{xz} = .00$. Additional variables may be obtained from the author by request.

REFERENCES

[1] Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.
[2] Hoffman, P. J. Predetermination of test weights. *Psychometrika,* 1958, **23,** 85-92.

Manuscript received 1/2/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON THE TRYON-KAISER SOLUTION FOR THE COMMUNALITIES

HENRY F. KAISER

UNIVERSITY OF ILLINOIS

The Tryon-Kaiser solution for the communalities is reviewed. Numerical investigation suggests that the procedure is applicable if and only if the correlation matrix has unique minimum rank communalities. This implies that this approach to the communality problem is not general enough to be of practical use.

In his recent review of the notion of communality from a cluster-analytic viewpoint, Tryon derives a formula "for the exact value of h^2" ([6], eq. 21). This treatment is interesting theoretically because it does not explicitly consider the dimensionality of the common-factor space. It is interesting practically because—at least for one example—it succeeded in obtaining an exact solution for the communalities.

Simultaneously and essentially independently, Kaiser [3] from a traditional factor-analytic point of view developed what ultimately is the same approach. His treatment consists of a derivation and an attempt to use the equations

$$\bar{h}_j^2 = \hat{h}_j^2 - 1/\hat{r}^{jj} \qquad (j = 1, 2, \cdots, n),$$

to provide an iterative solution for the communalities, where $\hat{h}_j^2$ is a trial value for the jth communality, $\bar{h}_j^2$ is a new (and hopefully improved) approximation to a solution, $\hat{r}^{jj}$ is the jth diagonal element of the inverse of the reduced correlation matrix with $\hat{h}_j^2$ in the diagonal, and n is the number of observed variables. This formula is applied as an attempt to compute the squared multiple correlation of test j on the remaining arbitrarily large number of tests in the hypothetical domain of content under consideration, a value which under very general conditions may be shown to equal the communality [2].

This note reports results obtained when this method of solving for the communalities was applied more extensively. In addition to Tryon's successful example, exact communalities for three further matrices were easily solved, using Kaiser's equations. These three examples, like Tryon's, had the property that the number of common factors was less than half the number of tests, as the off-diagonal elements of the correlation matrices involved had been generated artificially by multiplying an arbitrary factor matrix with this property by its transpose. For a second group of examples—six correlation matrices based on empirical data—this method did not yield communalities; the iterative procedure failed to converge.

The obvious question is whether this method will succeed only for artificial correlation matrices and not for data from the real world. The answer probably lies in a theorem of Albert's and in some results of Ledermann. If r is the number of common factors, Albert [1] proved that when $r < n/2$, there exist *unique* communalities such that the resulting reduced correlation matrix has rank r. If it is postulated that the Tryon-Kaiser procedure will be applicable if the correlation matrix has unique minimum rank communalities, Albert's theorem would account for the success with artificial matrices. On the other hand, *empirical correlation matrices do not have unique minimum rank communalities*. This follows from Ledermann [4]. He has shown that if

$$r \geqq \tfrac{1}{2}(2n + 1 - \sqrt{8n + 1}),$$

the communalities will *not* be unique. He has also shown that if r is to be less than $\frac{1}{2}(2n + 1 - \sqrt{8n + 1})$, special conditions must hold exactly among the off-diagonal elements of the correlation matrix. Because sample correlation coefficients are continuous random variables, Ledermann's special conditions can only hold with probability zero, and consequently in practice, unique communalities may occur only with zero probability.

By systematically varying r/n in constructing additional artificial correlation matrices, extensive further numerical investigation uniformly confirmed the hypothesis that the Tryon-Kaiser solution for the communalities will converge if and only if the correlation matrix under consideration has unique minimum rank communalities. Indeed, unique negative "communalities" from non-Gramian matrices (generated with imaginary factors) were easily found. Attempts to prove the hypothesis algebraically with methods described by Scarborough ([5], pp. 209–211) have not been successful.

The difficulty with the Tryon-Kaiser solution is that it is incomplete. What is needed is a criterion for selecting among the inevitable multiple solutions for empirical correlation matrices. What this criterion might be seems a difficult scientific (not mathematical or statistical) problem. It does not appear to have been explored systematically.

REFERENCES

[1] Albert, A. A. The minimum rank of a correlation matrix. *Proc. nat. Acad. Sci.*, 1944, **30**, 144-148.

[2] Guttman, L. "Best possible" systematic estimates of communalities. *Psychometrika*, 1956, 21, 273-285.

[3] Kaiser, H. F. Solution for the communalities: a preliminary report. Rep. No. 5, Contract No. AF 41(657)-76, Univ. California, 1956.

[4] Ledermann, W. On the rank of the reduced correlational matrix in multiple factor analysis. *Psychometrika*, 1937, **2**, 85-93.

[5] Scarborough, J. B. *Numerical mathematical analysis.* (2nd ed.) Baltimore: Johns Hopkins, 1950.

[6] Tryon, R. C. Communality of a variable: formulation by cluster analysis. *Psychometrika,* 1957, 22, 241-260.

Manuscript received 1/7/58

Revised manuscript received 3/9/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON THE USE OF TRIADS FOR PAIRED COMPARISONS

R. E. SCHUCKER

PURDUE UNIVERSITY

When scaling a large number of stimuli from comparative judgments, considerable savings in time and labor may be realized if stimuli are presented in triad form rather than in pairs. If, for N stimuli, the proper configuration of triads can be assembled so that all possible pairs appear once, the paired judgment matrix may be reproduced with one-third fewer judgments and two-thirds fewer presentations than would be required with complete pairing. A simple procedure is described for enumerating triad configurations for which N is an odd multiple of three.

Application of the traditional method of paired comparisons rapidly becomes unwieldy as the number of stimuli increases beyond 20. However, the labor required in eliciting and analyzing large numbers of paired judgments has been greatly reduced with the advent of punched card procedures [3]. In addition, the burden on the subject has been eased through the development of a partial pairing technique [5, 6]. With partial pairing, the total number of pairs to be judged may be reduced any desired amount by pairing each stimulus with fewer than the $N - 1$ remaining stimuli.

Where the paired comparison method is applied for the purpose of developing an interval scale, the investigator may desire that all stimuli be scaled with equal accuracy. That is to say, it is required that each stimulus be compared *equally often* with all *remaining stimuli*. It is readily apparent that the partial pairing technique does not fulfill the requirements for balance with respect to the estimation of stimulus scale values. Hence, when large numbers of stimuli are to be scaled, an alternative to partial pairing is needed if scale values are to be estimated with equal accuracy and the volume of paired judgments kept within reasonable bounds.

One approach in reducing the total number of pairs in paired comparisons has been through the use of triads. The format of the Kuder Preference Record is a familiar example of the grouping of stimuli in three's rather than in pairs [4]. If for the triad A, B, C, judgments of the type "most" and "least" are obtained, with respect to the ordering of the objects along a psychological dimension, preferences for the pairs (A, B), (A, C) and (B, C) may be recovered. Similarly, if for N objects the proper configuration of triads be assembled, so that every possible pair appears exactly once, information for the complete, paired judgment matrix is obtained with only one-third as many triads and two-thirds as many judgments as would be required using pairs and traditional full pairing. For example, complete pairing for an N of 57 requires 1596 pairs, while only $1596/3 = 532$ appropriate triads furnish the same paired information.

The problem remains of finding the triads needed for a given N in order that each element be paired once with every remaining element. The present paper touches upon a geometric solution that produces for certain N's a basic group of triads, from which the required configuration may be enumerated by cyclic permutation. Where the stimuli to be scaled are verbal in nature, the cycling procedure is readily adaptable to punched card equipment for the preparation of the triads and analysis of data.

Cyclic Enumeration of Configurations

In general, a configuration consists of N elements arranged in b sets of k elements each, with each element occurring in r sets, and each pair of elements occurring together in a set exactly λ times. The following relationships must hold.

$$\lambda = \frac{r(k-1)}{N-1}, \tag{1}$$

$$Nr = bk. \tag{2}$$

For the case under consideration $\lambda = 1$ and $k = 3$. Substituting these values in (1) and solving for r, yields $r = (N-1)/2$. Thus a configuration of triads satisfying the required relationships is possible only when N is odd. In addition, it should be pointed out that the conditions specified in (1) and (2) are necessary but not sufficient for the existence of a configuration. Having determined that a particular N satisfies the requirements, the task of assembling the triads of the configuration still remains.

Various methods for constructing configurations have been reviewed by Cox [2]. Class $\lambda = 1$ configurations have received considerable attention in the literature and numerous solutions are available for the subclass $k = 3$, where N is an odd multiple of three. Solutions for N of this form may be obtained by cycling basic groups of triads. Depending upon N, the enumeration of the complete configuration from the basic groups proceeds by 1-step cycles, 2-step cycles, 3-step cycles, or combinations of cycles. An example of enumeration by 1-step cycles is provided by the configuration for $N = 9$. From equations (1) and (2) it is seen that $r = 4$, $b = 12$. The configuration is completed by setting $N = 9$.

	Rep I		Rep II		Rep III		Rep IV	
row	set		set		set		set	
1	(1)	1 N 5	(4)	2 N 6	(7)	3 N 7	(10)	4 N 8
2	(2)	2 3 8	(5)	3 4 1	(8)	4 5 2	(11)	5 6 3
3	(3)	6 7 4	(6)	7 8 5	(9)	8 1 6	(12)	1 2 7

The triads are grouped into complete replications of the 9 elements. It will be observed that the remaining $r - 1$ triads in any row may be obtained by a succession of one-step cyclic permutations on the elements of the triad appearing in the same row in Rep I. For example, the addition of 1 to each element in triad (3) gives triad (6). Similarly, by cycling the elements of triad (6), triad (9) is obtained, and so forth. Note that the second element of each of the row-1 triads is not cycled and that for all triads one re-cycles to 1 on the next step when an element equals $N - 1$. Since the complete configuration can be generated from Rep I, the sets of Rep I may be arbitrarily designated as solution triads.

A second configuration is shown to illustrate enumeration by two-step cycles. For this example, $N = 15$, $r = 7$, and $b = 35$.

set	Rep I		
(1)	1	N	8
(2)	4	5	7
(3)	13	14	9
(4)	3	6	11
(5)	10	12	2

set	Rep II		
(6)	3	N	10
(7)	6	7	9
(8)	1	2	11
(9)	5	8	13
(10)	12	14	4

set	Rep III		
(11)	5	N	12
(12)	8	9	11
(13)	3	4	13
(14)	7	10	1
(15)	14	2	6

set	Rep IV		
(16)	7	N	14
(17)	10	11	13
(18)	5	6	1
(19)	9	12	3
(20)	2	4	8

set	Rep V		
(21)	9	N	2
(22)	12	13	1
(23)	7	8	3
(24)	11	14	5
(25)	4	6	10

set	Rep VI		
(26)	11	N	4
(27)	14	1	3
(28)	9	10	5
(29)	13	2	7
(30)	6	8	12

set	Rep VII		
(31)	13	N	6
(32)	2	3	5
(33)	11	12	7
(34)	1	4	9
(35)	8	10	14

Setting $N = 15$, the configuration satisfies equations (1) and (2). As before, the $r - 1$ remaining triads in any row can be generated starting with the corresponding triad of Rep I. For this example, however, cycling proceeds by 2-step increments. Thus, by adding 2 to each element of triad (2) one obtains triad (7), which gives rise in succession to triads (12), (17), (22), (27) and (32). Note that a re-cycling to 1 or 2 occurs on the next step for elements which equal $N - 2$ or $N - 1$.

Ball discusses variations on a geometric method for obtaining solution triads in configurations for which N is an odd multiple of three ([1], ch. 10). Essentially, the method involves the use of a circle with inscribed triangles, the points of which represent triad elements. The triangles of one replication are determined empirically, with the restriction that, when rotated within the circle, they generate the remaining replications of the configuration. Ball gives solution triads for $N = 9, 15, 21, 27, 33, 39, 45, 51, 57, 63, 69, 75, 81, 87, 93$ and 99.

Machine Cycling of Triads

The cycling procedures demonstrated in connection with the configurations for N's of 9 and 15 are readily adaptable to various card punch calculators, which can be utilized for preparing the triads on IBM cards as well as for generating configurations. Judgments can be recorded directly on the cards and the analysis facilitated by using tabulating equipment. The program requirements for enumerating triad configurations are relatively straightforward and may be deduced directly from the two illustrative examples.* Either modular arithmetic or the conventional arithmetic shown may be employed, although use of the former results in a shorter, more general program capable of handling cyclic steps of any magnitude.

REFERENCES

[1] Ball, W. W. R. Revised by Coxeter, H. S. M. *Mathematical recreations and essays.* (11th ed.) London: Macmillan, 1939.

[2] Cox, G. M. Enumeration and construction of balanced incomplete block configurations. *Ann. math. Statist.*, 1940, **11**, 72–85.

[3] Kephart, N. C. and Oliver, J. E. A punched card procedure for use with the method of paired comparison. *J. appl. Psychol.*, 1952, **36**, 47–48.

[4] Kuder, F. G. *Examiner manual for the Kuder Preference Record.* Chicago: Sci. Res. Associates, 1953.

[5] McCormick, E. J. and Bachus, J. A. Paired comparison ratings. II: The reliability of ratings based on partial pairings. *J. appl. Psychol.*, 1952, **36**, 188–192.

[6] Oliver, J. E. A punched card procedure for use with partial pairing. *J. appl. Psychol.*, 1953, **37**, 129–130.

Manuscript received 8/1/58

Revised manuscript received 10/16/58

---

*A generalized cycling program for the IBM 604 Electronic Calculating Punch has been deposited as Document number 5840 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

BOOK REVIEWS

DONALD DAVIDSON, PATRICK SUPPES, AND SIDNEY SIEGEL. *Decision Making, An Experimental Approach.* Stanford: Stanford University Press, 1957. Pp. 121.

The basic problem which the authors have set out to tackle in the studies reported in this book is the separation of the effects of psychological probability and utility in decision making. They have considered this problem both in its theoretical framework and in the experimental verification of the theories set forth. The first chapter gives an introductory discussion of the problems of empirical interpretation of theories of decision making under uncertainty. The second chapter fills almost half of the book and deals with the basic model proposed by the authors. The third chapter reports an experiment which was designed to measure the cardinal utility of nonmonetary outcomes and to use the computed utilities to predict further choices. Two models were compared, one a linear programming model and the other an ordinal model based on straightforward comparisons. The linear programming model turned out to be considerably superior to the other and both were much superior to a random guessing method; moreover, if thresholds are ignored in order to obtain a larger number of predictions the accuracy remains significantly better than chance. The fourth chapter considers the problem of formulating utilities for incomparable outcomes; in contrast with the two preceding chapters the considerations here are entirely axiomatic in character.

The second chapter offers an explicit theory for the explanation of individual decision making under conditions of risk, and reports an experiment designed to test the theory in certain limited situations. The first step in the theory is to construct an event which has a psychological probability of one-half. Next, a set of six outcomes is constructed so as to be equally spaced in utility, and from these a utility function is constructed which is adequate to account for a certain class of preference and indifference relations. The experimental results lead the authors to conclude among other things that (1) the theory provides a practical approach to the problem of resolution of utility and psychological probability in situations involving risk; (2) under suitably controlled conditions certain people make choices among risky alternatives as though they were attempting to maximize expected utility, and (3) for such persons it is possible to construct a utility function unique up to a linear transformation. The point of departure for this model was the work of Mosteller and Nogee (An experimental measurement of utility, *Journal of Political Economy*, 1951, **59**, 371–404); connections with other previous work are traced and a short but well-chosen bibliography is included.

In my opinion this book will take a prominent place in the literature of decision making, but it is also clear that it does not represent the final word on any of the points considered. The expository level of the book is excellent; the prospective reader should be prepared for a reasonable amount of mathematical development of the axiom-definition-theorem type.

R. M. THRALL

University of Michigan

WARREN S. TORGERSON. *Theory and Methods of Scaling.* New York: John Wiley and Sons, 1958. Pp. xiii + 460.

In 1950 the Social Science Research Council appointed a Committee on Scaling Theory and Methods to review the status of scaling procedures in the social sciences. This committee came to the inevitable conclusion that a good survey of the recent and prolific work on scaling procedures was necessary. In 1951, Warren Torgerson, as a Research Associate of the Council, undertook the preparation of a monograph on scaling procedures.

After seven years, two of which constituted a long lost weekend in the Navy, the monograph had become a book and was published.

The result of Torgerson's and the Committee's efforts is a book which will be of considerable influence and of great value to social scientists. It is an excellent summary of the state of the art and theory of measurement in the social sciences, the sort of book of which there is a very real shortage in psychology and the social sciences in general. It was not written as an undergraduate textbook, and probably cannot be used as such. While written primarily as a reference book for technical workers, it does seem possible to use it at the graduate level if a reasonable mathematical background is assumed on the part of the student. I suspect it will be more widely used than was anticipated.

Whatever shortcomings the book has are due more to the state of the art than to incorrect handling of the issues by the author. Torgerson provides an excellent organization of the scaling methods as they are in fact used; the various methods are covered in sufficient detail to enable anybody to use a particular method after a thorough reading of the appropriate section of the book. While extensive mathematical treatment of the various methods is provided, the author does not remain solely at the abstract mathematical level, but introduces the reader to the realities of collecting data and treating them as required by a particular scaling technique. The book is, in other words, a happy combination of erudite sophistication and down-to-earth realism.

The book starts out with the usual introductory and organizing chapters—in this case three of them. The first two chapters cover the nature of measurement, types of measurement, etc. The third chapter organizes measurement as it occurs in psychology into three classes: (1) where subjects are scaled; (2) where the stimuli are scaled, and subject differences are attributed to sampling error; and (3) where both stimuli and subjects are scaled from the same set of data. The first type has not led to any important scaling developments and is largely ignored. The second type, called the *judgment approach*, and the third, called the *response approach*, form the basis for the organization of the rest of the book. The term *response* for the last approach is a little unfortunate, since it does not differentiate that approach from the second, in which responses are also made.

The next seven chapters deal with the judgment methods. Three chapters are devoted to subjective estimates, fractionation, and equisection methods—those methods which involve some appreciation on the part of the subject of numerical values on the subjective continuum under consideration. The next four chapters deal with the discriminative, or differential sensitivity methods—all those methods which are based primarily on the Thurstone models. The last three chapters are concerned with the response methods. One chapter is concerned primarily with the Guttman techniques, one is concerned with Lazarsfeld's latent-structure model, and the last is concerned with the techniques developed by Coombs and his students for dealing with comparative response data.

There is no need to go into the specific content of each chapter. Each deals with its subject matter in a detailed and thorough way. The only chapter which I wish had not been included is that which introduces the differential sensitivity methods, and treats briefly and lightly of the traditional psychophysical methods. In the rest of the book if Torgerson covered a subject at all, he covered it thoroughly. In this one chapter, however, the coverage of the psychophysical procedures is entirely inadequate, and it would have been more in keeping with the rest of the book not to discuss the subject at all.

My major over-all reactions to the book were concerned less with what Torgerson wrote than with the state of measurement theory and practice in psychology today. For example, Torgerson sets up quite clearly the different types of data matrix which are used in scaling work, and differentiates methods on the basis of the nature of the data. He makes clear the kinds of assumptions which are made with regard to both the stimulus and the subject variables in such matrices, and thus provides a more fundamental look at the over-all picture than is customary. However, I found a desire to go back a step further, and to note that all basic sets of data involve three variables—stimulus, subject, and *response*—and that all of them have certain relations to the underlying true continuum. Just as each stimulus, or each subject, can be located on the continuum, so can each response—and there can be interactions between all three variables. It is not really necessary to assume anything fixed by the response, even a comparative response, and a truly general model for measurement would include solutions for the response as well as for stimuli and subjects. It is usual, of course, to have stimuli and subjects be orthogonal in experiments, while neither will normally be orthogonal to the response variable. It is quite possible, however, to give each subject a response and then tell him to find the stimulus which satisfies this response—just as is done with some fractionation procedures and with the equisection procedure. Any truly fundamental model for measurement must be able to deal with the relations between all three variables and the underlying continuum.

Actually, of course, we run into the reality of the lack of degrees of freedom for complete solutions, as Torgerson so often points out. And in fact the practical solutions with which we must deal often are such that an equivalent solution could have been obtained with simplifying assumptions other than those actually made. Using assumptions about the values or distributions of any of the three basic variables, we can solve for values of the other one or two. A change in the variables about which the assumptions are made will change the solutions as well. In other words, we cannot create more knowledge than the data give us; we can simply assign the knowledge to different variables.

In this frame of mind, I have one last reaction to report. It is that we have more sophistication about the nature of mathematical models of measurement than we do about experimental techniques for validating the models. Elegant scales can be constructed, but only after we have made enough assumptions to reduce the number of parameters to the number of available independent observations. Any verification of the assumptions requires the availability of more degrees of freedom, and experimental techniques must be devised which provide these degrees of freedom in a form appropriate to the assumptions being made. The mathematical models tell us only what *can* be so; better experimental techniques are necessary to tell us whether it *is* so.

These reactions are not intended as criticisms of the book, but rather as compliments to it. The book presents the whole range of material in sufficiently compact form that one is forced to try to get an overview. This fact, plus the over-all excellence of the presentation, will stimulate new and good research. I am tempted to say—and so I will say—that this book is a milestone.

WENDELL R. GARNER

The Johns Hopkins University

D. A. S. FRASER. *Nonparametric Methods in Statistics.* New York: John Wiley and Sons, 1957. Pp. x + 299.

Nonparametric Methods in Statistics (*NMS*) is an advanced work in statistical theory. *NMS* consists of two parts: an introduction to recent developments in the Neyman-Pearson tradition in statistical inference (Chapters 1 and 2) and an application of these developments to nonparametric statistics (Chapters 3, 4, 5, and 7). In addition, Chapter 6 is a survey of limit theorems useful in nonparametric theory.

The mathematical background necessary for comfortably reading *NMS* is a year's course in function theory. With less than advanced calculus the statements of many of the definitions and theorems are hardly intelligible. Mood (*Introduction to the Theory of Statistics*, McGraw-Hill, 1950) or preferably Cramér (*Mathematical Methods of Statistics*, Princeton University Press, 1946) are reasonable prerequisites in statistical theory. Without this background, *NMS* can possibly be used as a reference book for the theory. Siegel's book (*Nonparametric Statistics*, McGraw-Hill, 1956), at the other extreme, is nearly devoid of theory and mathematical content.

Besides the standard Neyman-Pearson optimum properties (e.g., most powerful, unbiased, and consistent), sufficiency, invariance, and completeness are stressed. These ideas are developed extensively and used in finding good nonparametric procedures—tests of hypotheses, point estimates, and tolerance intervals. Sufficiency and invariance have a strong intuitive appeal as criteria for optimality. Completeness is a mathematical condition that is useful when available. To illustrate these ideas consider the following problem. It is assumed that $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ are mutually independent random variables. Assume all of the X's have a common distribution and all of the Y's have a common distribution (not necessarily the same as that of the X's). How should one estimate Pr $(X_i < Y_j)$? (Pr $(X_i < Y_j)$ appears in the study of the Wilcoxon two-sample procedure.)

It is clear that the (temporal) order in which the observations are made is irrelevant and attention can be restricted to $X_{(1)}, \cdots, X_{(m)}, Y_{(1)}, \cdots, Y_{(n)}$, where $X_{(1)}$ is the smallest of $X_1, \cdots, X_m$; $X_{(2)}$ is the second smallest, etc. In short, the order statistics form a sufficient statistic for the problem. The parameter of interest, Pr $(X_i < X_j)$, will have the same value whether the original random variables are used or whether any monotone increasing function (e.g., exponential of the random variable) is used. Therefore it is reasonable to restrict attention to those estimators that will not change when an arbitrary monotone transformation of the observations is applied. This is the principle of invariance. For this problem, invariance implies that the estimator must be a function of the ranks. Completeness implies that there is a unique unbiased estimator (and hence minimal variance unbiased estimator) which is a function of the invariant sufficient statistic (the ranks). It is the number of pairs (X_i, Y_j), where $X_i < Y_j$, divided by mn.

The approach of the above paragraph is applied to many estimating and testing experiments which arise in practice, e.g., (a) making inferences from a random sample about the location parameter of a distribution; (b) testing the null hypothesis that c samples come from the same distribution against the alternative that the c samples come from distributions differing in location only; (c) making inferences about the amount of dependence in a bivariate distribution function, i.e., testing for independence and estimating correlation; and (d) constructing tolerance sets from univariate and multivariate data.

The orientation towards Neyman-Pearson theory and linear models (analysis of variance, etc.) explains the lack of emphasis on tests of goodness of fit. In keeping with the theoretical orientation, many mathematical examples are given, and, on the other hand, applied examples and tables of distributions are not given.

University of Minnesota I. Richard Savage

Philip J. McCarthy. *Introduction to Statistical Reasoning*. New York: McGraw-Hill, 1957. Pp. xiii + 402.

The author states his aims clearly in the Preface: ". . . a one-semester, nonmathematical course in statistics in which the instructor wishes to present a careful introduction to statistical reasoning. . . . A first course should emphasize the concepts of statistical reasoning rather than attempt to cover the wide variety of techniques. . . . Illustrative material should be drawn from investigations that are as significant as possible, and . . . has been chosen to range broadly over the social sciences. A very brief account of research problems usually accompanies each illustrative example and any student . . . may expect to improve his insight into the problems of research methodology in the social sciences. . . . This selection of illustrative material from the social sciences has also influenced to some extent the topics discussed in the book."

I agree with these aims and judge that the author has met them very well indeed; hence, I recommend this excellent book for a one-semester, nonmathematical, introductory statistical course in the social sciences. It excels in the choice of exercises and illustrations drawn from important research publications in the social sciences, has many good examples and fine figures, tables, and charts to illustrate the important problems.

How does this book differ from some of its better competitors? First, it emphasizes, as the title states, statistical reasoning and inference rather than statistical techniques and manipulation. Rather than trying to provide a reference book on a large variety of statistical techniques, the author concentrates on a thorough presentation of the nature of statistical inference. He does this briefly in two chapters which contain the two most useful statistical methods for social scientists: Chapter 8, The Binomial Probability Model and Statistical Inference; and Chapter 9, Drawing Statistical Inferences from the Arithmetic Mean of a Large Sample.

Second, the statistical problems and examples are not alienated from their origins and their destinations. The problems and examples are established firmly in the substantive problems from which they arise, and in the problems of collection and processing which precede the data. The relation of sample to population is often and well developed. The meaning of statistical tools and of statistical inference is constantly emphasized; that statistics and probability statements are guides to action and to decisions is the spirit that pervades the presentation (although there is no formal presentation of statistical decision function theory). Chapter 2, The Components of Statistical Investigation, presents this approach early and well.

Third, the writing is rigorous and precise. The presentation is not mathematical and does not require a mathematical background, but it does demand close attention and careful reading. The style is clear, precise, and to the point, and will aid the student through the hard thinking that rigorous statistical inference demands.

Chapters 3 through 7 deal well with the necessary introductory topics of distributions, their locations and spread, and the elements of probability. Chapter 10 gives in 30 pages an excellent, lucid, and penetrating presentation of Elements of Sample Design. Each of the last two chapters presents an important technique for social scientists: Chapter 11, Chi-Square Procedures for Qualitative Data, and Chapter 12, The Linear Association of Two Quantitative Variables.

Some instructors will prefer to add one or two additional techniques to complete the course; for economists, perhaps time series and indexes; for psychologists, perhaps the difference of two correlated means and the elements of experimental design. The student will have to read his assignments thoroughly and sometimes repeatedly. The instructor will have to explain the finer points at greater length. But perhaps that is not too much to ask of a scientific subject in a scientific age. I think this is a good book.

LESLIE KISH

University of Michigan

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

AN APPROACH TO MENTAL TEST THEORY*

FREDERIC M. LORD†

EDUCATIONAL TESTING SERVICE

It currently seems to me that the heart of mental test theory is the concept of true score.

The trouble starts as soon as it is realized that the actual test score obtained by a particular individual might just as well have been some numerical value other than the one actually observed. The individual examinee might have guessed differently, might have been less nervous, or might have slept better the night before. The testing conditions might have been different—lighting, facilities for use of paper and pencil, presence or absence of distraction. Finally, any one of many different but equally appropriate psychological tests might have been constructed and administered.

Mental test theory must deal with all these kinds of disturbing influences. However, we are not really interested in each of the different test scores that an examinee might obtain under all sorts of conditions. We are interested in something approximated by these scores. This something may be called the true score on the test.

Many definitions of true score are possible. Each is open to serious objections. The hope is that when a particular definition of true score is used, this definition will lead to useful approximate statements about the true value that we are interested in, even though this is not adequately defined.

Let us assume that the error of measurement, defined as the difference between true score and observed score, has an expected value of zero in all circumstances. This is the only assumption needed in order to obtain accurate estimates of the mean true score for a group of examinees, provided that the number of examinees is fairly large. I mention this in order to point out that some useful and meaningful results about true score can be obtained without making a lot of highly questionable assumptions.

What further assumptions or definitions can profitably be used to make additional inferences about true scores? I plan to consider three different types of approaches, which I will first describe briefly.

Perhaps the simplest and most straightforward approach is to assume that one can obtain independently at least two, and preferably more, test scores such that the scores for any single examinee differ only because of errors of measurement. From a practical point of view, it is reasonable to make such an assumption about the carefully matched split-halves of a test, or even split-thirds, fourths, fifths, etc., provided the original test contains a sufficient number of items.

---

*Presidential address delivered to the Psychometric Society, Cincinnati, Ohio, September 8, 1959.

†Much of the research discussed here was carried out under Contracts Nonr-2214 (00) and -2752(00) with the Office of Naval Research, Department of the Navy.

Instead of this, a second approach goes back to the separate items of which the test is composed. According to one line of reasoning, a very large pool of items is defined in terms of the properties of the items actually at hand; true score is then defined in terms of the hypothetical performance of the individual examinee on this very large pool of items, assuming no practice effect, etc. A closely related model simply assumes the items at hand are a random sample drawn from a very large pool.

Finally, a somewhat different approach is to make some restrictive assumptions about the frequency distribution of the errors of measurement. The assumption that this distribution is normal with constant variance is adequate for many common situations. A more plausible model assumes that the distribution of the errors of measurement is binomial.

It seems to me that all three of these approaches to the problem of making inferences about true scores are very useful. They lead to different models for a theory of test scores. Five of these will be discussed in some detail, both practical and theoretical, skipping in part or in whole those results that are already well known.

Before proceeding, it will be worthwhile to point out the distinction between, on the one hand, the true score of a test and, on the other hand, a variable variously called the common factor of the test items, "the ability underlying the test," or the latent continuum. A superficial distinction is that the true score usually has a limited range of possible values—in the most usual case, from zero up to some maximum possible score such as the total number of items in the test; whereas the "ability underlying the test" is conveniently taken as having a possible range from $-\infty$ to $+\infty$.

A more basic distinction is that nonparallel tests of the same psychological dimension will each necessarily have a different true-score metric just as each has a different observed-score metric. On the other hand, all tests of the same psychological dimension will have the same underlying-ability metric, once this has been properly defined. In this distinction lies the main virtue of the latent continuum—it makes possible rigorous comparisons between nonparallel tests of the same psychological dimension (for example, comparisons to determine which of two nonparallel tests is more discriminating in a specified subrange of the latent continuum). There is, of course, a perfect curvilinear correlation between the true score on a test and its latent continuum; the latent continuum is simply a nonlinear transformation of the true-score scale. Thus the true scores on two nonparallel tests of the same psychological dimension have a perfect curvilinear correlation—one is simply a nonlinear transformation of the other.

The advantages of the latent continuum are achieved at the cost of the more restrictive assumptions required for a stronger mathematical model. If we do not need to compare nonparallel tests of the same psychological dimension (in some cases even when we do), the stronger assumptions of the latent continuum model can be avoided and a consideration of true scores will be quite adequate.

Before proceeding with a discussion of the results obtainable from various mathematical models for true score, it may be pointed out that certain of these results may be applicable to the most varied types of problems in all scientific areas—typically, whenever the problem involves a large number of observed values, each of which contains a sizable unbiased error of measurement. A familiar example in the area of psychometrics is a problem where one has a large number of item-test correlations or item-criterion correlations. Since each of these is subject to a large sampling error, various questions arise. One would like to know the frequency distribution of the "true" item indices—the values that would be found if sampling errors could be eliminated. One would like to have an estimate of the true correlation for each individual item. One would like to know the effect of discarding all items whose observed correlations are below some fixed value. Given the sampling distribution of the individual correlation coefficient, these questions and related ones could be answered, at least approximately, by using certain of the approaches to be discussed here.

Matched-Forms Model

The matched-forms model for the relation between true score and observed score involves only two, very weak assumptions: (i) the expected value of the error of measurement is always zero (this statement applies to any individual or group of individuals selected in a fashion that is experimentally independent of the scores containing the errors in question), (ii) the true score of each individual is assumed to be the same on each of the test forms. Given these assumptions, no further definition of true score is needed. The matched forms themselves provide the necessary definition of their true score.

Under these assumptions, given k matched forms of a test, estimates can be obtained for all the moments of the frequency distribution of true scores up through the kth order. From three matched forms, an estimate of the skewness of the true-score distribution can be obtained; from four matched forms, an estimate of kurtosis; and so forth.

The kind of results obtained under this model may be illustrated by the simple relationship in (1), which holds exactly in the population of examinees and approximately in large samples of examinees.

$$\sum_a X_{1a} X_{2a} \cdots X_{ka} = \sum_a \xi_a^k , \tag{1}$$

i.e., the kth moment of the true scores is equal to the product of the examinee's raw scores on k matched forms of the test, this product being averaged over all examinees.

The moments of the frequency distribution of the errors of measurement can also be estimated, as can the multivariate moments involving both true score and error of measurement. From available formulas, all these estimates will be unbiased in random samples of examinees, even when the number of examinees is small, and regardless of the shape of the frequency distribution sampled.

What can be done with these estimated moments? Methods for setting bounds on a distribution function from its first k moments are available [1, 12, 22, 28, 37]. Let me quote from M. G. Kendall ([13], p. 83): "For all ordinary purposes, therefore, a knowledge of the moments, when they all exist, is equivalent to a knowledge of the distribution function: equivalent, that is, in the sense that it should be possible *theoretically* to exhibit all the properties of the distribution in terms of the moments $\cdots$.

"If now two distributions have moments up to order n equal they must have the same least-squares approximation [by a polynomial of degree n] $\cdots$ thus distributions which have a finite number of lower moments in common will, in a sense, be approximations one to another. We shall encounter many cases where, although we cannot determine a distribution function explicitly, we may ascertain its moments at least up to some order; and hence we shall be able to approximate to the distribution by finding another distribution of known form which has the same lower moments. In practice, approximations of this kind often turn out to be remarkably good, even when only the first three or four moments are equated."

In practice, a Pearson curve having the same first four moments as the true scores may be used to approximate the true-score distribution. The Charlier and Edgeworth series are available when more than four moments are known. Other asymptotic approximations are summarized in a recent article by Wallace [34].

The mathematical derivations for the matched-forms true-score model and the necessary formulas for estimating the moments are scheduled to appear in [20]. A discussion of the implications of the model has appeared in [19]. In order to avoid undue repetition here, I will merely remark that the estimated moments of the errors of measurement and the estimated bivariate moments between true score and error for any given set of data can be used to make a significance test of the hypothesis that the errors of measurement are distributed normally and independently of true score. Certain empirical results from such significance tests will be discussed at an appropriate later point.

The assumption that has been made here of strict parallelism between the test forms can be relaxed if one is willing to assume that the true scores on all the forms are linearly related to each other. With this assumption, one has the prototypical problem of linear structural relationship. Selected references in this very relevant area are [2, 9, 10, 11, 14, 15, 21, 24, 26, 33, 34, 35, 36].

Rationally Equivalent Forms Model

The fact that the matched-forms model is conveniently applied in practice to split-thirds, split-fourths, or split-fifths of a test suggests that there should be a true-score model that bears the same relationship to the matched-forms model as the Kuder-Richardson reliability coefficient bears to the split-half method of computing reliabilities. Such a method can easily be formulated. It may be called the rationally equivalent forms model.

For present purposes, the Kuder-Richardson approach may be thought of as follows. Imagine a large number of hypothetical forms of the test, all rigorously parallel to each other and to the actual test at hand. (By "rigorously parallel" is meant that in any sufficiently large preselected group of examinees, the group statistics computed from the test scores will be the same, no matter which form of the test is used.) The crux of the Kuder-Richardson approach is to estimate the correlation between any two parallel forms of the test from item statistics computed for the single form actually available. In order to achieve this, one further assumption is needed, which, it seems to me, cannot be classified as part of the assumption of rational equivalence, as some writers have done. The algebraic derivation turns up with a term representing the covariance between item i in the test actually at hand and the corresponding item in the hypothetical test. Some assumption is required. The usual assumption is, of course, that the average value of this unknown covariance is the same as the average interitem covariance for all the items in the test at hand.

With this assumption, the correlation between rationally equivalent forms of the test can now be estimated. Since this correlation is a reliability coefficient, it can be used to estimate the variance of true scores and the variance of the errors of measurement. It is implicit in all this that true score on the test at hand is to be defined as the average score that would be obtained on a very large number of rationally equivalent forms. It is only a short step from this line of reasoning to point out that the item statistics on the one test actually at hand can be used not only to estimate the variance but also the higher moments of the true-score distribution.

Let X_{ua} be the score of examinee a on test u. Suppose that this score is the sum of the examinee's scores on n items, so that

$$X_{ua} = \sum_{i=1}^{n} x_{uia} , \tag{2}$$

where x_{uia} is the score of examinee a on the ith item in test u. The desired results are obtained by first substituting (2) into (1). For example, if there are three parallel tests, then

$$\sum_a \xi_a^3 = \sum_{g=1}^{n} \sum_{h=1}^{n} \sum_{i=1}^{n} \Big(\sum_a x_{1ga} x_{2ha} x_{3ia} \Big). \tag{3}$$

Application of the line of reasoning just outlined now transforms (3) into

$$\frac{1}{N} \sum_a \xi_a^3 = \frac{1}{n(n-1)(n-2)} \sum \sum_{g \neq h \neq i} \sum \frac{1}{N} \sum_a x_{1ga} x_{1ha} x_{1ia} \, . \tag{4}$$

It is thus seen that under the assumptions made the third moment of the true scores for sufficiently large samples of examinees can be computed from the available item statistics on a single form of the test. Similar formulas may be written for the higher true-score moments. Furthermore, quantities like those on the right of (4) can be expressed as a linear function of the moments of the observed-score distribution, to be denoted by m'. Thus (4) can be written

$$\frac{1}{N} \sum_a \xi_a^3 = \frac{1}{n(n-1)(n-2)} (m_3' - 3m_2' + 2m_1'^2). \tag{5}$$

Under this mathematical model for true scores, given a sufficiently large sample of examinees, there is no place for the usual process of statistical estimation. Once the assumptions are granted, the population of test items is defined exactly by the item statistics in the tests actually at hand. I am not completely happy with this model. If tests are to be rationally equivalent, it seems clear that corresponding items in different tests will usually be more alike than would two different items chosen at random. This is certainly not enough reason for discarding the model, however.

The model to be discussed next starts from a somewhat different set of premises, which, however, lead to formulas not quite identical to those illustrated by (4).

The Item-Sampling Model

This is the same model that was previously called the matrix-sampling model [16]. In this model, it is assumed that the items in the test at hand can be considered as a random sample of items from a very large pool or population of items. The true score of an individual examinee is here conveniently defined as a proportion—as the probability that an item chosen at random from the pool of items will be one that he will answer successfully. More or less standard techniques of statistical inference permit the estimation of the examinee's true score, thus defined, from his responses to the items at hand. Under this model, given that the number of examinees, N, is sufficiently large so that we are willing to neglect quantities of order $1/N$, the observable quantity on the right-hand side of equation (4) is no longer equal to the third true-score moment but is an unbiased estimate of it in random samples of test items. The accuracy of this estimate depends on n, the number of items in the test at hand.

There is another important difference between the results obtained under the item-sampling model and those obtained under the rationally equivalent forms model. Under the rationally equivalent forms model, if we wish to know the numerical value of a product or power of true-score moments, it is only necessary to multiply together the numerical values computed for the moments in question. Under the item-sampling model, this is not sufficient. A product of unbiased estimates of moments is not the same as an unbiased estimate of the product of moments. Special formulas must be derived to provide unbiased estimates of the various products called for by different formulas. The necessary formulas are available [16, 17] so that unbiased estimates can be computed for any true-score moment about the origin or about the mean up through the fourth order.

Also, formulas are presented for the bivariate moments between true score and observed score. This makes possible the estimation of the regression of true score on observed score, even when this is curvilinear. By this method, therefore, one can, at least in theory, predict the true score of a given examinee from his observed score without assuming a linear regression.

A useful result is obtained when linear regression is assumed. The formula for the regression coefficient in this case turns out to be very similar to the Kuder-Richardson formula-21 reliability coefficient. In fact, it differs from this coefficient only by quantities of order n^{-2}.

Now, a little algebra makes it apparent that this regression coefficient should in fact also be a reliability coefficient. In fact, there seems to be more intuitive meaning if one defines reliability as the regression of true score on observed score rather than as a ratio of two variances. I was at first somewhat puzzled, therefore, to find discrepancies between the regression coefficient obtained under this model and the Kuder-Richardson formula-21 reliability. The answer to this puzzle has been independently provided by Nageswari Rajaratnam, working with Lee Cronbach and, more recently, but probably not completely independently, by myself. Rajaratnam defines reliability in terms of the ratio of two variances and then derives a small-sample formula for estimating reliability in the type of situation where the item-sampling model holds. When the number of examinees, N, is large, as explicitly assumed in my formula, but not in hers, the two formulas are the same.

Equation (6) gives one form of the large-sample formula for this new reliability coefficient, which is also the slope of the regression of true score on observed score,

$$\hat{\rho} = 1 - \frac{\bar{p}\bar{q} - s_z^2}{(n-1)s_z^2 + s_p^2}, \tag{6}$$

where $\bar{p}$ is the average item difficulty, $\bar{q}$ is $1 - \bar{p}$, s_z^2 is the variance of proportion-correct observed scores, and s_p^2 is the variance of item difficulty.

The relationship of this reliability coefficient to K-R (21) is not very apparent when the formula is presented in this form. The difference between the two coefficients can be shown to be

$$\hat{\rho} - r_{21} = \frac{s_p^2(1 - \hat{\rho})}{(n-1)s_z^2} = O\left(\frac{1}{n^2}\right). \tag{7}$$

As long as the test has more than zero reliability, the difference between the two coefficients is actually a quantity of order n^{-2}, since the quantity $1 - \hat{\rho}$ in the numerator is itself a quantity of order n^{-1}.

When the number of examinees is large, the sampling fluctuations in $\hat{\rho}$ or r_{21}, due to sampling of items, are of the order of $n^{-3/2}$. Thus the difference between $\hat{\rho}$ and r_{21} is a half order of magnitude smaller than the sampling fluctuations in either. The difference between them will be negligible in typical testing situations. This comparison differs from that between the K-R (20) and K-R (21) coefficients, since their difference is of order n^{-1} and is thus larger than the sampling error of either. From the point of view of the item-sampling model, the new coefficient $\hat{\rho}$ (or better, the small sample version derived by Rajaratnam) seems preferable to r_{21} from a logical point of view even though their difference is usually small compared to their sampling fluctuations.

In working with various mathematical models for true scores, it is often disturbingly difficult to find ways of checking theoretical deductions from the models against observable experimental results. It is therefore of interest to bring forward two practical applications of the item-sampling model. In these as yet unpublished studies, the true-score model is used to predict the moments of observed-score distributions. These latter distributions are subsequently obtained and the actual moments compared with the predicted values.

In the first study, the problem is to predict the effect of lengthening a test upon the shape of the frequency distribution of observed scores. For example, if we have a skewed distribution of observed scores, will doubling the length of the test tend to reduce the skewness or to increase it? It turns out that either result may occur, depending upon the nature of the test, and also upon the nature of the group of examinees tested.

On the one hand, one can imagine a group of examinees whose proportion-correct true scores are symmetrically distributed over a very small range, say, from .85 to .95. As the test consists of items that are rather easy for this group, the observed scores will tend to have a negatively skewed distribution. If the test is made longer and longer, however, the observed-score distribution will approach the symmetry of the true-score distribution. In this case, therefore, lengthening the test decreases the skewness of the observed-score distribution.

As an example of the opposite situation, consider a case where the group of examinees tested has a somewhat skewed true-score distribution in the narrow range between .45 and .55. If the test is short and unreliable and is composed of 50-percent difficulty items, the symmetrical distributions of the errors of measurement will almost completely swamp the skewness of the true-score distribution, so that the observed-score distribution on a short test will appear nearly symmetrical with observed proportion-correct scores, possibly ranging from zero to one. In this case, lengthening the test will increase the skewness of the observed-score distribution.

What is needed is a formula for determining from the available item statistics and score statistics on a single form of the test which of these things will happen—whether the skewness will decrease or increase as the test is lengthened. Similar questions, of course, may be asked for the kurtosis of the distribution.

How is it that true-score theory can be of use here? The reasoning is simply as follows. It is given that the longer test is parallel to the shorter test except for length, i.e., the proportion-correct true score for any examinee will be the same on either test. Formulas are available for unbiased estimates of true-score moments in terms of item statistics and score statistics on the short form of the test. Such formulas are also available in terms of statistics on the long form of the test. We have, in principle, two unbiased estimates of each moment of the true-score distribution, one from each form. If these two unbiased estimates are equated, a set of formulas expressing statistics on the longer test in terms of statistics on the shorter test is obtained. For example, if the long-form statistics are distinguished by a prime, (4) leads to

$$\text{(8)} \qquad \frac{1}{n'(n'-1)(n'-2)} \sum_{g' \neq h' \neq i'}^{n'(n'-1)(n'-2)} \sum \sum x_{g'a} x_{h'a} x_{i'a} \doteq \frac{1}{n(n-1)(n-2)} \sum_{g \neq h \neq i}^{n(n-1)(n-2)} \sum \sum x_{ga} x_{ha} x_{ia} \, .$$

Since a moment of the longer test is a linear function of quantities like those on the left of (8), the moments on the longer test can be estimated from the item statistics on the shorter [17]. As already mentioned, algebraic manipulation makes the computations involved in equations like (8) feasible. This procedure has been applied empirically; satisfactory agreement between predicted and actual values was found for the data studied.

In the second study, the problem was to predict the bivariate moments of the scatterplot between two parallel forms of a test from the item statistics of either form alone. Formulas for this have been developed [17] by a line of reasoning exactly parallel to that just described. These bivariate moments determine the shape of the scatterplot and, in particular, the shape of the regression of one form on the other (this regression need not be strictly linear even though the forms are parallel*). This procedure was also applied empirically and satisfactory agreement was found between predicted and actual values.

Gaussian Errors of Measurement

A somewhat different perspective emerges if one is willing to make more specific assumptions about the frequency distribution of the errors of measurement. It is frequently an adequate approximation to assume that the errors of measurement are normally distributed with zero mean and constant, experimentally determinable variance. For convenience, this true-score model will be referred to as the Gaussian error model. Some extremely interesting and useful results can be deduced from this model, many of which I am sure have never been applied in mental test theory.

First of all, as noted in ([16], pp. 2–3), when this model holds, all the cumulants of the true-score distribution except the second are exactly equal to the cumulants of the observed-score distribution. Thus, if the variance of the errors of measurement can be estimated, then any moment of the true-score distribution can be expressed as a polynomial in the moments of the observed-score distribution.

In my recent work I have been searching for methods of dealing with true scores without computing moments. The reason for doing this is that the use of moments has certain practical disadvantages, which I should mention briefly. In the first place, third- and fourth-order moments are subject to large sampling fluctuations, and estimated third- and fourth-order true-score moments even more so. Even with 2,000 examinees, the standard error of the estimated fourth-order moment is uncomfortably large. Higher order moments will seldom be useful. If only four moments of a distribution can be used, there is no reason to believe that this is a particularly efficient method of approximating the distribution. The use of Pearson curves is probably often effective in practice, but their effectiveness cannot be guaranteed. It is well known that the Charlier series and the Edgeworth series may give poor fits to skewed distributions when only four moments are used. These difficulties motivate a search for methods of making inferences about true scores without computing the moments of the observed-score distribution.

Before proceeding, it is appropriate to ask whether every frequency distribution of observed scores is compatible with the Gaussian error model.

---

*Ferguson has shown that, with improbable exceptions, this regression will not be linear unless the true score is normally distributed in the group tested [7].

It must be admitted first of all that this model can never be strictly appropriate for describing an observed frequency distribution of number-correct or proportion-correct scores, simply because such scores represent a discrete and bounded variable whereas the scores produced under the Gaussian error model are necessarily continuous and unbounded (except, of course, in the degenerate case when the variance of the errors is zero; this case will be ruled out of further consideration). Avoiding a strict interpretation of the Gaussian error model, suppose we replace the histogram of the number-correct scores by a smooth frequency curve that approximates it, running from minus infinity to plus infinity. Will the model always be compatible with such a smooth frequency curve? It is intuitively clear that given some fixed value for the variance of the errors of measurement, there must be some limit to the sharpness of curvature in the frequency distribution of observed scores. The presence of errors of measurement necessarily obscures in the observed-score distribution any sharp detail that might exist in the true-score distribution. There is therefore a limit to the sharpness of detail that can appear in the observed-score distribution under this model. Necessary and sufficient conditions for the frequency distribution of observed scores to be compatible with the Gaussian error model are given by Pollard [25] and by Standish [29]. In the present context, their conclusions are probably primarily of theoretical rather than practical interest.

The Gaussian error model is completely represented by (9):

$$f(x) = \int_{-\infty}^{\infty} g(\xi)N(x - \xi)\, d\xi, \tag{9}$$

where $f(x)$ is the frequency distribution of observed scores, $g(\xi)$ is the unknown frequency distribution of true scores, and $N(x - \xi)$ is the normal distribution for the variable $e = x - \xi$ having zero mean and known standard deviation, σ. It is helpful to visualize this equation as referring to a scatterplot between true score on the abscissa and observed score on the ordinate. The quantity under the integral sign is the product of the frequency distribution of true scores and the conditional frequency distribution of observed score for a given true score. This product is equal to the bivariate distribution between true score and observed score; for fixed values of x and ξ, the quantity under the integral sign may be thought of as representing the frequency in any cell of the scatterplot. The integral sign represents a set of summations. The equation says simply that the frequency of occurrence of a particular score x is equal to the quantity obtained by summing all the cell frequencies in the corresponding row of the scatterplot.

Equation (9) is an *integral equation*. The quantity $f(x)$ is known as the *convolution* of $g(\xi)$ and $N(\xi)$. The mathematical problem is typically to estimate $g(\xi)$ when $f(x)$ is known. More specifically, (9) is a Fredholm equation of the first kind. At the moment, it seems to me that methods for solving such integral equations ([3], ch. 14; [4]; [31], sec. 3.15; and [5]) will be an important aid in making inferences about true scores. Trumpler and Weaver ([32], chs. 1.4 and 1.5) give at least five different methods for solving equations such as (9). In addition, they go on to more advanced problems that are of great interest here, such as "Correction of a Bivariate Distribution for Observational Errors" and "Statistical Determination of the Functional Relationship between Variables."

Almost all methods for solving (9) start off by replacing the continuous variable ξ by a discontinuous variable. For example, suppose that the true-score distribution $g(\xi)$ can be approximated by a discontinuous distribution denoted by $G(\xi)$. Equation (9) becomes

$$f(x_m) = \sum_r G(\xi_r)N(x_m - \xi_r), \tag{10}$$

where x_m is any specified value of the observed score, and ξ_r is one of a set of specified values for the true score. For fixed values of x and ξ, the function $N(x_m - \xi_r)$ is a tabled value of the normal curve ordinate. Equation (10) is a linear equation in the unknown frequencies, the G's. As long as these unknown frequencies are no more numerous than the known frequencies, the f's, (10) can be solved for the G's, giving an approximate representation of the frequency distribution of true scores.

I am anxious to point out the essential simplicity of (10). If the number of equations is equal to the number of unknowns, then the solution merely involves computing the inverse of the matrix whose elements are the values of $N(x - \xi)$. Moreover, once it has been computed, this inverse can be used over and over again for all sorts of different tests containing the same number of items. Each unknown frequency in the true-score distribution is thus expressed as a weighted average of the frequencies in the observed-score distribution.

While emphasizing a certain simplicity about the problem, I should also mention that there are certain serious complications, caused by the fact that the observed-score distribution will never fit (9) exactly, if only because of sampling fluctuations. This may pose serious obstacles in the way of making valid inferences about the true-score distribution. The procedure just outlined is likely to give too good a fit when applied to actual data, with the result that some of the G's may turn out to be negative, or that the frequency distribution represented by the G's may have a very irregular shape. An important current problem is finding procedures that will deal with these types of difficulty.

In a recent paper, Gaffey [8] considers from the point of view of statistical inference a method devised by Sir Arthur Eddington, the astronomer, to solve equation (9). Gaffey obtains formulas for statistical estimators of the G's and also for the sampling variance of these estimators.

I think you will be interested in seeing how simple Gaffey's method is. It will be convenient to describe the method in a practical application, although the method itself is perfectly general. The data used were the observed scores of 388,071 students on a 50-item test. These data were chosen because of the large number of cases and because the score distribution looked approximately normal. Only .0002 of the cases made perfect or zero scores. The procedure used was to group the frequency distribution into 26 class intervals and then to compute the first, second, third, and fourth differences of the column of frequencies. Under the Gaussian-error model, any true-score frequency can then be estimated by starting with the corresponding observed-score frequency, subtracting from it one-fourth of the corresponding second difference, and adding one thirty-second of the corresponding fourth difference.

I have gone into detail on this method in order to indicate what a very simple procedure may be found for estimating a true-score distribution directly in terms of the frequencies of observed scores. Unfortunately, the results obtained by the application of Gaffey's method to the data described show that these data do not meet the assumptions of the method. The estimated true-score frequencies formed a bell-shaped distribution and appeared at first to be satisfactory, but the variance of the estimated distribution was found to be much too large in view of the known variance of the errors of measurement in the data. Dr. Gaffey was kind enough to correspond with me on this problem. It appears likely that the discrete and bounded character of a frequency distribution of number-correct scores cannot be reconciled with the assumptions underlying the Gaussian error model within a sufficient tolerance to allow the application of Gaffey's methods. It is to be hoped that the binomial model, to be discussed next, will avoid some of these difficulties.

Before leaving the Gaussian error model, let me describe briefly a beautiful result obtained by Eddington and summarized by Trumpler and Weaver ([32], pp. 128–131). Eddington derives the regression of true score on observed score in the form

$$M_{\xi\cdot x} = x + \sigma^2 \frac{f'(x)}{f(x)}, \tag{11}$$

where the term on the left is the mean true score for a given observed score, σ^2 is the variance of the errors of measurement, and $f'(x)$ is the first derivative of the frequency distribution of the observed scores. This equation is based on no approximations and involves no assumptions other than the basic assumption of the Gaussian error model.

How effective (11) will be for estimating true scores from number-correct observed scores remains to be determined. Numerical methods are certainly available for approximating the first derivative from the first and higher differences of the observed-score distribution. In any case, results such as (11) give encouragement to efforts to find simple, direct methods for estimating true score from observed score without assuming a linear regression.

Although it is clear that the Gaussian error model cannot fit number-correct score data exactly, it is of interest that data have been found for which the Gaussian-error model is rejected by appropriate tests of statistical significance. As already mentioned, the necessary significance tests were derived from the very general matched-forms model, which I discussed first of all. The significance tests were applied to the vocabulary scores in four widely different groups of one thousand examinees each. The Gaussian-error hypothesis was rejected for each of the four groups at the .01 significance level. The results will be reported in [18]. These results do not mean that the Gaussian-error model should be rejected from mental test theory, but they do justify an attempt to find a more appropriate model for estimating true scores.

The Binomial Error Model

In the Gaussian error model, the observed scores have a normal distribution for any fixed value of true score. In trying to improve on this model, it is natural to consider the effect of assuming the observed scores to have a binomial, instead of normal, distribution when true score is fixed. This assumption is equivalent to the integral equation

$$f(x) = \int_0^1 g(\zeta)B(\zeta, x)\, d\zeta, \qquad (x = 0, 1, \cdots, n), \tag{12}$$

where ζ is the proportion-correct true score

$$\zeta = \xi/n, \tag{13}$$

and where $B(\zeta, x)$ is the binomial distribution

$$B(\zeta, x) = \binom{n}{x}\zeta^x(1 - \zeta)^{n-x}. \tag{14}$$

This model is intended to be appropriate for dealing with number-correct scores, not for scores obtained by formula scoring or other methods. Whatever its faults, there seems to be no question but that the binomial error model is a better approximation than is the Gaussian error model. It is also in many ways a fairly simple model to handle mathematically.

For one thing, it is not difficult to express the moments of the true-score distribution in terms of the moments of the observed-score distribution:

$$\mu_r'(\zeta) = \frac{(n - r)!}{n!}\, m_{[r]}'(x), \qquad (r = 0, 1, \cdots, n). \tag{15}$$

The left side of (15) represents the rth moment of the true-score distribution. The quantity $m_{[r]}'(x)$ designates the rth factorial moment of the observed-score distribution. A factorial moment ([13], pp. 56–60) is simply a linear function of the regular moments, m'_r ; for example,

$$m'_{[4]}(x) = m'_4 - 6m'_3 + 11m'_2 - 6m'_1 \; . \tag{16}$$

It turns out that (15) is the same as the set of equations illustrated by (5), representing the rationally equivalent forms model. It is thus seen that the binomial error model is completely equivalent to the rationally equivalent forms model, insofar as the latter is concerned with predicting the shape of the frequency distribution of true scores from that of observed scores. The two models are not interchangeable, however, for purposes of inferring such things as the scatterplot between true and observed scores.

For some time now, I have been trying out various numerical methods for using equation (12) to infer the frequency distribution of true scores from a given distribution of observed scores. One set of data may be of interest here. They are the scores of more than 2000 professional people on a 30-item figure matrix test. An interesting question arises because a full one-sixth of the examinees have scored at the chance level or below—a full 10 percent score below chance. Only 7 percent of the examinees get as many as half the items right after correction for guessing. The presence of so much guessing in the score distribution must tend to obscure the shape of the true distribution of competence in the group tested; so it will be of interest to examine the distribution of true scores. A still more interesting question is whether or not the scores below the chance level simply represent random deviations from a true score at the chance level, or whether some of the examinees actually have true scores that are below the chance level. I already have enough data clearly to answer the second question. There is no doubt but that a substantial number of examinees have true scores below the chance level. This result could occur in one of two ways, both of which seem to deserve further investigation. On the one hand, it may be that some examinees do not understand the directions to the test and systematically proceed in the wrong fashion. The second, and I think more likely possibility, is that many of the distractors for the test items are so ingeniously contrived as to be more attractive to some people than are the correct answers.

In order to apply a binomial error model to data of this kind where there are many omitted items, it is necessary for the statistician to supply a response for each item that each examinee has omitted. In the case of five-choice items, the result is the same as would have been obtained if the examinee had tossed a five-sided die to determine which response to make on the omitted items. While certain questions can be raised as to the legitimacy of this procedure, it seems that it is not likely to give rise to true scores below the chance level unless these were already present in the data.

Whenever the frequency distribution of true scores has been inferred from a given set of data by means of an integral equation such as (9) or (12), it is always possible to use the integral equation to reconstruct the bivariate frequency distribution of true scores and observed scores, since this bivariate frequency distribution is simply the integrand on the right side of the equation. Given the bivariate frequency distribution between observed score and true score for any given set of data, it is a simple matter to obtain by the numerical methods the mean true score corresponding to each different observed score—in other words, the regression of true score on observed score.

I have been making a search for more direct methods of estimating this regression—for a method similar, perhaps, to Eddington's method in the Gaussian error case. There is an important theoretical difference between the Gaussian case and the binomial case in this connection. Under the Gaussian model, if a given $f(x)$ is compatible with the model, it is likely to determine completely the frequency distribution of true scores. Under the binomial model, the observed data consist of $n + 1$ frequencies. These determine at best only n moments of the true-score distribution, by means of (15). The higher moments of the true-score distribution can, within narrow limits, be assigned arbitrarily without impairing the ability of the true-score distribution to reproduce the observed data. All of these true-score distributions with identical lower order moments will look very much alike since they all have the same approximating polynomial of degree n; however, the indeterminacy gives rise to certain mathematical problems, as will be seen in a moment.

A simple recurrence relationship, apparently as yet unpublished, can be derived from (12):

$$M_{\zeta\cdot x} = \frac{x+1}{n-x}\,\frac{f(x+1)}{f(x)}\,(1 - M_{\zeta\cdot x+1}). \tag{17}$$

In words, the mean true score for any given observed score, x, can be deduced from the frequency of occurrence of the score, the frequency of occurrence of the next larger score, $x + 1$, and the mean true score at $x + 1$. Thus, if the mean true score could be inferred for any value of x between 0 and n, the recurrence relationship would give the mean true score at all other values of x.

As a result of the indeterminacy of the true-score distribution, however, there is no precise way of getting started. We have n independent linear equations, but we have $n + 1$ unknown values of $M_{\zeta\cdot x}$ to determine. There seem to be some rather good practical expedients for dealing with this indeterminacy, however. For one thing, it is wholly reasonable to require that the relationship between x and the estimated value of $M_{\zeta\cdot x}$ be monotonic increasing, at least in that part of the score range where the most data are available—that is, near the mode of the score distribution. This requirement immediately sets rather narrow limits to the values that can be assumed by

$M_{\zeta.x}$. Figure 1 shows two different curves obtained for $M_{\zeta.x}$ by choosing the values of $M_{\zeta.11}$ first as large as possible and then as small as possible consistent with monotonicity in the region $x = 10, 11, 12$. The figure suggests that for these data the average true score can be determined within a fairly narrow range for any given observed score.

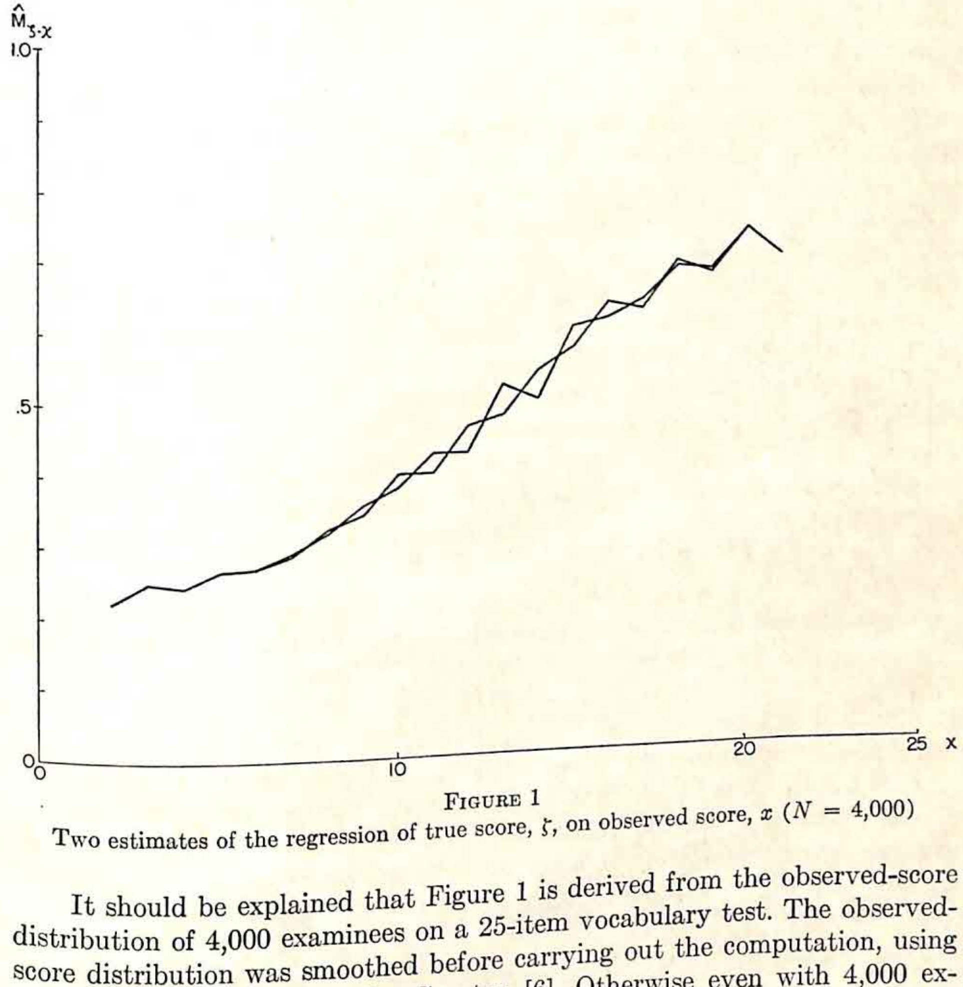


FIGURE 1
Two estimates of the regression of true score, ζ, on observed score, x ($N = 4{,}000$)

It should be explained that Figure 1 is derived from the observed-score distribution of 4,000 examinees on a 25-item vocabulary test. The observed-score distribution was smoothed before carrying out the computation, using a five-point formula given by Cureton [6]. Otherwise even with 4,000 examinees, the regression of true score on observed score would have been excessively irregular, due to chance fluctuations in the adjacent frequencies of the observed-score distribution.

The jagged line in Figure 2 is an estimate of the regression of true score on observed score for a 50-item quantitative test. In this case, there were 388,071 examinees. The observed-score distribution was not smoothed in obtaining Figure 2. The figure is presented to show the excellent agreement

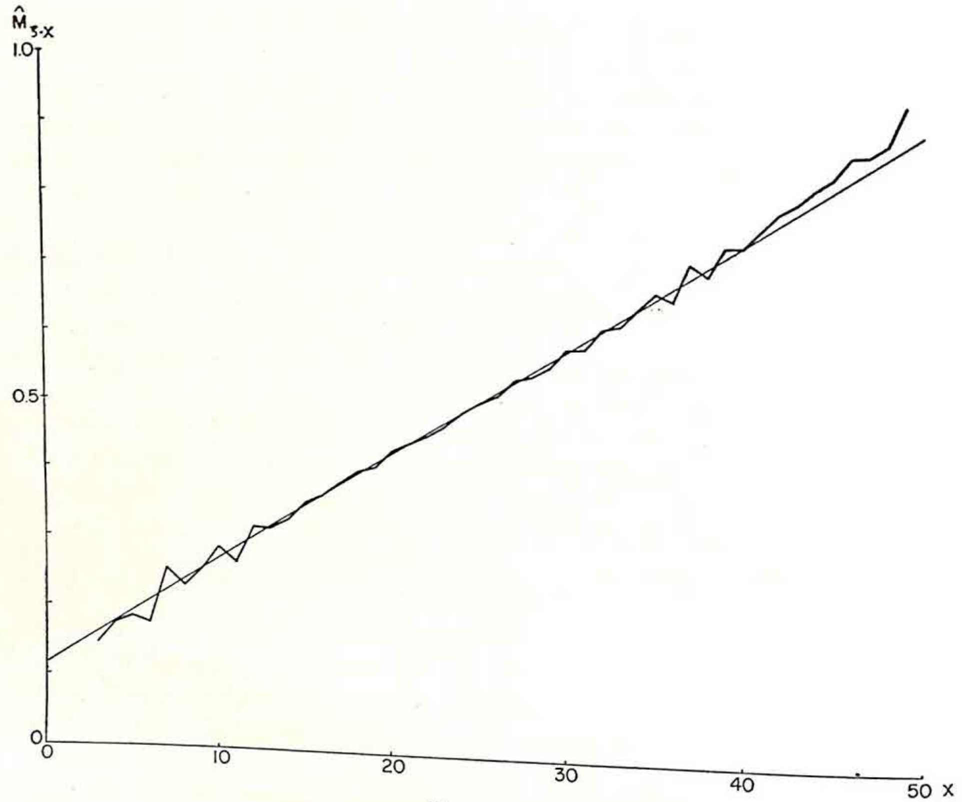


FIGURE 2

Regression of true score on observed score estimated from K-R (21) (straight line) and from binomial error model (N = 388,071)

between the results obtained from (17) and the linear regression of true score on observed score computed from the K-R (21) reliability coefficient, represented by the straight line in the figure. It turns out that under the binomial error model, the linear regression coefficient for predicting true from observed score, given a large number of examinees, is the K-R (21) coefficient, not the K-R (20) coefficient or the new coefficient given in (5), above.

In the case of Figure 2, there appears to be little need for using (17) since the regression is very close to linear and can be well represented by the classical equation. This result might have been guessed in advance from the fact that the observed-score distribution is very nearly normal.

The device used to obtain the jagged regression line in Figure 2 was to try various values of $M_{\xi.x}$ for values of x near the mode and to choose the value that made the regression line most nearly straight in the middle. This would seem to be quite a justifiable way of using (17). A more elaborate but quite feasible procedure would be to determine the values of $M_{\xi.x}$ so as to minimize the jaggedness of the regression line in some least squares sense. For other methods related to this problem, see Robbins [27] and Steinhaus [30].

Much more remains to be done in working out convenient numerical methods for dealing with the relatively simple problems already discussed. There are also certain more advanced problems, for which test theory should be able to provide an answer. It is hoped that the present methods may lead to solutions for some of the following, for example.

1. Given the observed-score distribution for a single group of examinees on each of two tests that are known to measure the same psychological dimension, but that are not otherwise parallel, estimate the possibly curvilinear functional relation between their true scores, and deduce from this the scatterplot of their observed scores.

2. Given, further, the observed-score distribution of one of these two tests in a second and somewhat different group of examinees, estimate the observed-score distribution of the other test in the second group of examinees. Also estimate the scatterplot for the second group of examinees.

3. Devise a method for determining whether two tests do or do not measure the same psychological dimension, i.e., a method for determining whether the true scores on two tests have a perfect curvilinear relationship or not.

In closing, I would say that all five of the true-score models outlined yield good results when applied appropriately. All five will probably continue to be of active interest. Empirical studies comparing the results obtained under different models would help to clarify their various advantages and disadvantages.

REFERENCES

[1] Barton D. E. and Dennis, K. E. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrics*, 1952, **39**, 425–426.

[2] Basmann, R. L. The computation of generalized classical estimates of coefficients in a structural equation. *Econometrica*, 1959, **27**, 72–81.

[3] Buckingham, R. A. *Numerical methods*. London: Pitman, 1957.

[4] Collatz, L. *Numerische Behandlung von Differential-Gleichungen*. Berlin: Springer, 1951.

[5] Crout, P. D. An application of polynomial approximation to the solution of integral equations arising in physical problems. *J. math. Phys.*, 1940, **19**, 34–92.

[6] Cureton, E. E. and Tukey, J. W. Smoothing frequency distributions, equating tests, and preparing norms. *Amer. Psychologist*, 1951, **6**, 404. (Abstract)

[7] Ferguson, T. On the existence of linear regression in linear structural relations. *Univ. Calif. Publications in Statistics*, 1955, **2**, No. 7.

[8] Gaffey, W. R. A consistent estimator of a component of a convolution. *Ann. math. Statist.*, 1959, **30**, 198–205.

[9] Geary, R. C. Inherent relations between random variables. *Proc. Roy. Irish Acad.*, 1942, **47**, 63–76.

[10] Geary, R. C. Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica*, 1949, **17**, 30–58.

[11] Geary, R. C. Relations between statistics: The general and the sampling problem when the samples are large. *Proc. Roy. Irish Acad.*, 1943, **49**, 177–196.

[12] Johnson, N. L. and Rogers, C. A. The moment problem for unimodal distributions. *Ann. math. Statist.*, 1951, **22**, 433–439.

[13] Kendall, M. G. *The advanced theory of statistics*. London: Griffin, 1948.

[14] Kendall, M. G. Regression, linear structure, and functional relationship. *Biometrika*, 1951, **38**, 11–25.

[15] Koopmans, T. C. and Reiersøl, O. The identification of structural characteristics. *Ann. math. Statist.*, 1950, **21**, 165–181.

[16] Lord, F. M. Statistical inferences about true scores. *Psychometrika*, 1959, **24**,1–17.

[17] Lord, F. M. Use of true-score theory to predict moments of univariate and bivariate observed-score distributions. Princeton: Educ. Test. Serv. Res. Bull. 59–6, 1959.

[18] Lord, F. M. An empirical study of the normality and independence of errors of measurement in test scores. *Psychometrika*, in press.

[19] Lord, F. M. Inferences about true scores from parallel test forms. *Educ. psychol. Measmt*, 1959, **19**, 331-336.

[20] Lord, F. M. The joint cumulants of true values and errors of measurement. *Ann. math. Statist.*, in press.

[21] Madansky, A. The fitting of straight lines when both variables are subject to error. *J. Amer. statist. Ass.*, 1959, **54**, 173–205.

[22] Mallows, C. L. Note on the moment-problem for unimodal distributions when one or both terminals are known. *Biometrics*, 1956, **43**, 224–227.

[23] Moran, P. A. P. A significance test for an unidentifiable relation. *J. Roy. statist. Soc.* B, 1956, **18**, 61–64.

[24] Neyman, J. and Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrica*, 1948, **16**, 1–32.

[25] Pollard, H. Distribution functions containing a Gaussian factor. *Proc. Amer. math. Soc.*, 1953, **4**, 578–582.

[26] Reiersøl, O. Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 1950, **18**, 375–389.

[27] Robbins, H. An empirical Bayes approach to statistics. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. Berkeley: Univ. California Press, 1956. Pp. 157–163.

[28] Royden, H. L. Bounds on a distribution function when its first n moments are given. *Ann. math. Statist.*, 1953, **24**, 361–376.

[29] Standish, C. N-dimensional distributions containing a normal component. *Ann. math. Statist.*, 1956, **27**, 1161–1165.

[30] Steinhaus, H. The problem of estimation. *Ann. math. Statist.*, 1957, **28**, 633–648.

[31] Tricomi, F. G. *Integral equations*. New York: Interscience, 1957.

[32] Trumpler, R. J. and Weaver, H. F. *Statistical astronomy*. Berkeley: Univ. California Press, 1953.

[33] Tukey, J. W. Components in regression. *Biometrics*, 1951, **7**, 33–69.

[34] Wallace, D. L. Asymptotic approximations to distributions. *Ann. math. Statist.* 1958,**29**, 635–654.

[35] Wolfowitz, J. Estimation of the components of stochastic structures. *Proc. Nat. Acad. Sci.*, 1954, **40**, 602–606.

[36] Williams, E. J. Significance tests for discriminant functions and linear functional relationships. *Biometrics*, 1955, **42**, 360–381.

[37] Zelen, M. Bounds on a distribution function that are functions of μ_0 to μ_4 . *J. Res. Nat. Bur. Standards*, 1954, **53**, No. 6, paper 2556.

Manuscript received 9/8/59

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

THEORETICAL MODELS OF CHOICE AND STRATEGY BEHAVIOR: STABLE STATE BEHAVIOR IN THE TWO-CHOICE UNCERTAIN OUTCOME SITUATION*

SIDNEY SIEGEL

THE PENNSYLVANIA STATE UNIVERSITY

A theoretical approach to the understanding of human behavior in uncertain outcome situations is suggested, an approach which draws upon utility theory, decision-making theory, and statistical association theory. Experimental evidence supporting this approach as opposed to alternative approaches is summarized. Three different formalizations are presented, and a variety of experimental tests is suggested.

The purposes of this paper are:

I. to discuss a controversy which has arisen between two theoretical approaches pertaining to the behavior of human beings in a situation involving choices under uncertainty;

II. to suggest a different approach, set in a decision-making and utility theory framework, to account for the choice behavior to which both the other theories have reference;

III. to present three mathematical models—each involving concepts from decision-making and utility theory—which make predictions about human behavior in the choice situation under discussion;

IV. to present and discuss some experimental evidence which supports the general features of these models.

The Controversy Between Two Theoretical Approaches to Choice Behavior

The predictions which people make when placed in a two-choice uncertain outcome situation have received considerable attention in recent years [2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 17, 18, 20, 21]. In the classical experimental situation, as first used by Humphreys [14], the subject is asked to predict whether a light will or will not appear after a signal stimulus. Following the signal, either of two mutually exclusive alternative events can occur: the light can appear or the bulb can remain unlit. The two events occur with fixed but unequal probabilities, say π_1 and π_2, in a random sequence for a number of trials. In recent years, two lights have often been used, one of which appears with probability π and the other with probability $1 - \pi$ on each trial. In some cases, events other than lights have been used. In any case, the events are independent of the subject's behavior—regardless of what he does, one of the two will occur on each trial. The subject is instructed to do his best to predict correctly which of the two events will occur. He is given no other verbal information about the situation, but he is allowed to witness the events and thus to determine for himself the correctness or incorrectness of each of his predictions.

---

*This paper was written while the author was at the Center for Advanced Study in the Behavioral Sciences, and it has benefited greatly from discussions with many colleagues there. In particular, some of the features of Models I and II emerged from work with Robert P. Abelson, and Model III was developed to its present state partly as a result of consultations with Claude Shannon, and John C. Harsanyi (of the Department of Economics at Stanford University). The treatment of all three models was sharpened during many discussions with Robert M. Solow.

There are two theoretical models of interest here which yield predictions about an individual's behavior in a situation of this sort. The first is Estes' statistical learning model [5], which yields the prediction that subjects will learn to match their response ratios (the relative frequencies with which they predict each of the two events) to the actual probabilities of occurrence of the events, π and $1 - \pi$. The same prediction is given by the Bush-Mosteller model [2] when certain restrictions are applied to the parameters of that model. A number of studies report findings in support of this prediction [8, 12, 13, 15].

The second model, formalized by von Neumann and Morgenstern [22], is a game-theoretic model. According to some interpreters, one prediction consistent with this model is that a person will learn to maximize the expected frequency of correct predictions. To do this, he will predict the more frequent event on all trials.

To contrast the predictions from the two models, consider, for example, a two-choice uncertain outcome situation in which over a series of trials the two events, E_1 and E_2, occur with probabilities π and $1 - \pi$ of .75 and .25, respectively. Estes' model asserts that a person will learn to make predictions which will tend toward stabilizing at prediction of E_1 on 75 percent of the trials ($p = .75$) and prediction of E_2 on 25 percent of the trials. A person who adopts this "strategy" will have .625 as his expected proportion of correct predictions, for

$$\begin{aligned} E_x &= p\pi + (1 - p)(1 - \pi), \\ &= .75(.75) + .25(.25) = .625. \end{aligned}$$

On the other hand, if the subject were to adopt the "pure strategy" of predicting E_1 on every trial ($p = 1.0$), he would have .75 as his expected proportion of correct predictions, for

$$1.0(.75) + 0(.25) = .75.$$

One straightforward prediction from the game-theoretic model, then, would seem to be that a person would behave in the second way. Evidence giving some support to this prediction may be found in [10, 11, 17, 18, 21].

Because of the considerations sketched above, some behavioral scientists who have been influenced by game-theoretic principles have asserted that people who match their response ratios to the probabilities of the events are acting irrationally, in that they are failing to maximize their expected proportion of correct predictions, a goal they could accomplish by predicting the more frequent event on every trial. The empirical fact is that in this situation most people do, after many trials, stabilize at matching their response ratios to the probabilities of E_1 and E_2 . To assert that this is irrational is to rely on a highly restrictive meaning of that term. It was pointed out by Bernoulli [1] in the first half of the eighteenth century that any theory of rational behavior which does not incorporate the concept of the utility or *subjective* value of the outcomes rather than their *objective* value will lead to paradoxes of the kind under discussion. As Simon has reminded us ([20], p. 271), one must bear in mind the distinction between objective rationality (rationality as viewed by the experimenter) and subjective rationality (behavior that is rational, given the perceptual and evaluational premises of the subject). The purpose of the following section is to suggest a resolution of the apparent paradox which has been described, a resolution which incorporates the utility theory approach.

A Decision-Making Approach

From decision-making theory [3, 4], a hypothesis of *maximization of expected utility* may be drawn which will account for both sorts of prediction behaviors—matching response ratios and maximizing the expected frequency of correct predictions. According to this approach, whether a person will tend toward one or the other prediction strategies depends on certain conditions related to the reinforcement inherent in the situation. Where *utility* is understood to refer to the subjective value of an outcome, the general hypothesis is that a person will behave as if he were attempting to maximize expected utility in any instance. The components of the total utility vary in magnitude from situation to situation and thus a person's strategy to maximize expected utility varies.

It is reasonable to suppose that when a person is in a situation in which the only payoff attached to the outcomes is the satisfaction of having his prediction confirmed by the event or the dissatisfaction of having his prediction disconfirmed by it, making a correct prediction of the rarer event has greater utility for the person than making a correct prediction of the more frequent event. The person derives satisfaction from playing a game with the machine, trying to outwit it. Moreover, there is the matter of monotony, both kinesthetic and cognitive: predicting the more frequent event on all trials would engender the monotony of pressing the same button (a common method of stating predictions) on trial after trial for hundreds of trials, in addition to the monotony of the same cognitive response (e.g., left light) on trial after trial. Under such circumstances, a subject may maximize expected utility by matching his response ratio to the actual probabilities of occurrence of the two events. For him, the cost of an incorrect prediction is very low, whereas the gain from this strategy in terms of other utilities, such as the utility of gambling and the utility of variability, may be relatively high. By choosing a mixed strategy (i. e., splitting his predictions in some proportion), such a subject may maximize his own total satisfaction. If such an account is correct, then a decision-making model and the Estes model could yield the same predictions concerning the stable state (asymptotic) behavior of a person in a two-choice no payoff situation.

However, if the utility attached to correct predictions is increased by a change in the conditions of the game, i.e., if there is potential satisfaction beyond that of having an event confirm one's prediction, or if the cost attached to an incorrect prediction is increased, then the prediction yielded by the hypothesis of maximization of expected utility diverges from the prediction yielded by the Estes model. That is, if the decision-making theory approach proposed here is correct, the introduction of systematic variation in the reinforcements (and thus utilities) attached to correct or incorrect predictions should be followed by systematic differences in subjects' responses. As the utility of a correct prediction is increased and/or the cost (negative utility) of an incorrect prediction is increased, the person's prediction of the more frequent event should tend to 100 percent, i.e., he may be expected to choose a pure strategy rather than a mixed one.

To make precise predictions about behavior, it is necessary to have more precise formulations of the decision-making theory approach than the informal comments above. The next section presents three such formulations. Each of these mathematical models is set in a decision-making framework and is based on considerations from utility theory. The models have emerged from an attempt to construct a unified theory to account for the stable-state choice behavior to which both the Estes model and the game-theoretic model have reference.

Alternative Utility Models of Choice Behavior

Of the three models which are to be presented, Model I is the simplest. Central to it are two concepts: the utility of a correct prediction, and the utility of variability—the utility of varying one's responses, or relieving monotony. In this model, there is no explicit consideration of the utility of gambling. Thus, there is an implicit assumption, and this may be a reasonable one, that when data on decision making consist of averages for groups of subjects, there is little, if any, systematic effect from the utility of gambling, since the utility of gambling is positive for some people and negative for others.

Model II is richer than Model I in that it includes the concepts of the first model—although slightly different mathematical properties are ascribed to them—and it also deals with the utility of gambling.

Model III contrasts with the first two in that it incorporates the Shannon information measure, and also in that it seems to lead to somewhat different predictions than are yielded by either Models I or II.

Model I

Let π = the probability of occurrence of the more frequent event,
p = the proportion of times the subject chooses the more frequent event,
a = the marginal utility of a correct prediction, and
b = the marginal utility of varying one's responses.

If the expectation that a subject's prediction will be correct, E_x , is

$$\begin{aligned} E_x &= [p\pi + (1 - p)(1 - \pi)] \\ &= [(1 - \pi) + p(2\pi - 1)], \end{aligned}$$

then the expected utility of a correct prediction, U_r , is

$$U_r = aE_x = a[(1 - \pi) + p(2\pi - 1)].$$

If U_v = the utility of varying one's responses = $f(p)$, such that U_v is symmetrical and maximal for $p = .5$, e.g., $p(1 - p)$, (which, though arbitrarily chosen, has considerable intuitive appeal), then

$$U_v = bp(1 - p).$$

The total expected utility of a particular strategy p is

$$\begin{aligned} U(p) &= U_r + U_v \\ &= a[(1 - \pi) + p(2\pi - 1)] + bp(1 - p). \end{aligned}$$

The strategy p which maximizes expected utility is at

$$\frac{dU(p)}{dp} = 0$$

and is

$$p = \frac{a(2\pi - 1)}{2b} + \frac{1}{2}.$$

If $\alpha = a/b$ then

$$p = \alpha(\pi - \tfrac{1}{2}) + \tfrac{1}{2}. \tag{1}$$

It may be seen from (1) that when and only when $a = b$ and thus $\alpha = 1$, $p = \pi$. That is, when the marginal utility of a correct prediction equals the marginal utility of varying one's responses, Model I makes the same prediction regarding asymptotes (stable state behavior) as the Estes model. This situation seems to hold when the only reward in the situation is knowing whether one's prediction is correct or not. However, if the reinforcements inherent in the situation are increased, say by adding monetary rewards and/or costs to the outcomes, and thus the utility of a correct prediction and/or the negative utility of an incorrect prediction increases, then $a > b$ and $\alpha > 1$. For such conditions, the predictions regarding asymptotes yielded by Model I diverge from the predictions yielded by the Estes model.

The predictions Model I yields concerning a subject's choice and strategy behavior may be generally stated as follows. For any π,

(i) $p = 1$ when $\alpha \geq \dfrac{1}{2\pi - 1}$,

(ii) $1 > p > \pi$ when $1 < \alpha < \dfrac{1}{2\pi - 1}$,

(iii) $p = \pi$ when $\alpha = 1$,
(iv) $p < \pi$ when $\alpha < 1$.

The third prediction, that $p = \pi$, is the prediction yielded by the Estes model, and is a special case of Model I. The first, second, and fourth predictions are ones which it would seem that the Estes model is not prepared to make.

In Model I, $\hat{\alpha}$ may be estimated from data from (1),

$$\hat{\alpha} = \frac{p - (1/2)}{\pi - (1/2)}.$$

In the final section of this paper, experimental results are reported which give some support to Model I. It will be seen that when *S*s are run in the Estes situation under three different reinforcement conditions, the values of $\hat{\alpha}$ which may be estimated from the data are:

No Payoff condition, $\hat{\alpha} = 1.00$;
Reward condition, $\hat{\alpha} = 1.44$;
Risk condition, $\hat{\alpha} = 1.80$.

A relatively simple test of Model I is being conducted by the author at present, in which the above values of $\hat{\alpha}$ are utilized under different values of π. Preliminary results strongly support Model I.

A more crucial test of Model I would entail the experimental manipulation of α by variation of b, the marginal utility of varying one's responses. In the Estes No Payoff condition, with an increase in b, one would predict $p < \pi$. With a decrease in b, one would predict $p > \pi$. S. W. Becker, in a personal communication, has described data he has collected at Stanford University confirming these predictions.

Model II

Let π = the probability of occurrence of the more frequent event, as in Model I,
p = the proportion of times the subject chooses the more frequent event, as in Model I,
a = the marginal utility of a correct prediction when and only when the subject chooses the more frequent event,
b = the marginal utility of a correct prediction when and only when the subject chooses the less frequent event, and
c = the marginal utility of varying one's responses.

The expected utility of a correct prediction, $E_z(U_r)$, is

$$E_z(U_r) = ap\pi + b(1 - p)(1 - \pi),$$

and the utility of varying one's responses, U_v, is $f(p)$, as in Model I, and is

$$U_v = cp(1 - p).$$

The total expected utility of a particular strategy p is

$$U(p) = ap\pi + b(1 - p)(1 - \pi) + cp(1 - p).$$

The strategy p which maximizes $U(p)$ is at

$$\frac{dU(p)}{dp} = 0,$$

and is

$$p = \frac{a\pi - b(1 - \pi) + c}{2c}.$$

If $\alpha = a/c$ and if $\beta = b/c$ then

$$p = \frac{(1 - \beta) + \pi(\alpha + \beta)}{2}. \tag{2}$$

It may be seen from (2) that when $a = b = c$, and thus $\alpha = \beta = 1$, Model II yields, as a special case, the Estes prediction that $p = \pi$. It will be shown later that the case when $\alpha = \beta = 1$ is not the only case for which $p = \pi$ according to Model II.

Model II is richer than Model I in several respects. One is that it makes possible experimental studies involving differential payoffs for the more frequent and less frequent events. In addition Model II may lead to the measurement of an individual's specific utility of gambling. The predictions Model II yields concerning a subject's choice and strategy behavior may be generally stated as follows. For any π:

(i) $p = 1$ when

$$\alpha \geq \frac{(1 - \pi)\beta + 1}{\pi} \tag{A$_1$}$$

and/or

$$\beta \leq \frac{\pi\alpha - 1}{1 - \pi}\,; \tag{B$_1$}$$

(ii) $1 > p > \pi$ when

$$\frac{(1 - \pi)\beta + 1}{\pi} > \alpha > \frac{(1 - \pi)\beta + (2\pi - 1)}{\pi}$$

and/or

$$\frac{\pi\alpha - 1}{1 - \pi} < \beta < \frac{\pi\alpha - (2\pi - 1)}{1 - \pi}\,;$$

(iii) $p = \pi$ when

$$\alpha = \frac{(1 - \pi)\beta + (2\pi - 1)}{\pi} \tag{A$_\pi$}$$

and/or

$$\beta = \frac{\pi\alpha - (2\pi - 1)}{1 - \pi}\,; \tag{B$_\pi$}$$

(iv) $p < \pi$ when

$$\alpha < \frac{(1 - \pi)\beta + (2\pi - 1)}{\pi}$$

and/or

$$\beta > \frac{\pi\alpha - (2\pi - 1)}{1 - \pi}.$$

The first, second, and fourth predictions—that p will be 1, between 1 and π, or less than π—are predictions that the Estes model is not prepared to make. The third prediction—that p will be at π—is the Estes prediction, and occurs here as a special case (or set) in Model II. It is interesting to note that the second set of predictions above is obtained by proving the inequalities $(A_1) > (A_\pi)$ and $(B_1) < (B_\pi)$.

In Model II, there are two parameters, α and β, which must be estimated before the value of p can be predicted:

$$p = \frac{(1 - \beta) + \pi(\alpha + \beta)}{2}. \tag{2}$$

To test Model II, two successive studies would be needed, the first to obtain estimates of the parameters. Subjects could be run under two π's and one level of reinforcement, R_1 . Data would be required, that is, from a study of the following design.

	π_1	π_2
R_1	p_1	p_2

With π_1 and π_2 set, and with p_1 and p_2 observed, according to Model II,

$$p_1 = \frac{(1 - \beta) + \pi_1(\alpha + \beta)}{2},$$

$$p_2 = \frac{(1 - \beta) + \pi_2(\alpha + \beta)}{2}.$$

Solving these two equations simultaneously, estimates of β and α are as follows:

$$\hat{\beta} = \frac{2(p_1\pi_2 - p_2\pi_1)}{\pi_1 - \pi_2} + 1,$$

$$\hat{\alpha} = \frac{2(p_1 - p_2)}{\pi_1 - \pi_2} - \hat{\beta}.$$

With $\hat{\beta}$ and $\hat{\alpha}$ computed, a second study could then be conducted using a π different from those above, say π_3 . Under the same reinforcement, R_1 , as was used above, a test of Model II would be a test of the prediction

$$p_3 = \frac{(1 - \hat{\beta}) + \pi_3(\hat{\alpha} + \hat{\beta})}{2}.$$

An important difference between Model I and Model II is that Model II contains separate expressions for the utility of correctly predicting the more frequent event and the utility of correctly predicting the less frequent event. Thus, whereas both models predict that p will depend on the reinforcement hinging on correct predictions, only Model II yields a prediction for the situation in which the reinforcement contingent on correctly predicting the more frequent event differs from the reinforcement contingent on correctly predicting the less frequent event. According to this model, by systematically varying these two types of reinforcement it should be possible to induce variations in p from zero to one!

With such an experimental set-up it should be possible to obtain a measure of a subject's specific utility of gambling. To obtain such a measure, it would be necessary to know the subject's utility of money, and this could be determined by the method developed by Davidson, Siegel, and Suppes [3]. The experimenter could then determine how large a change in money is necessary to change the subject's strategy, p, from some stable value to unity. This result, under the proper experimental conditions, could yield a measure, in monetary units, of the subject's specific utility of gambling.

Model III

The third model makes a potentially interesting connection between two theoretical approaches: utility theory and information theory.

Let π = the probability of occurrence of the more frequent event, as in Models I and II,

p = the proportion of times the subject chooses the more frequent event, as in Models I and II,

a = the marginal utility of a correct prediction when there is a payoff (say monetary) in addition to the satisfaction of knowing a prediction has been confirmed, and

b = the marginal utility of "reflecting" or "reproducing," in some manner (stochastic), the information in the event system.

The expectation of a correct prediction, E_z , is

$$E_z = [p\pi + (1 - p)(1 - \pi)] = [(1 - \pi) + p(2\pi - 1)].$$

Then the expected utility of a correct prediction with payoffs over and above the satisfaction of knowing a prediction has been confirmed, $E_z(U_m)$, is

$$E_z(U_m) = a[(1 - \pi) + p(2\pi - 1)].$$

The logarithm of the expectation that the subject will reflect the information, or stochastic structure, of the event system, for n trials, is

$$\log \binom{n}{\pi n} p^{\pi n}(1 - p)^{(1-\pi)n}.$$

Utilizing Stirling's formula and expanding, the logarithm of the expectation for any particular trial (i.e., $n = 1$) is

$$[\pi \log p + (1 - \pi) \log (1 - p)] - [\pi \log \pi + (1 - \pi) \log (1 - \pi)].$$

Thus the expected utility involved is

$$U_I = b\{[\pi \log p + (1 - \pi) \log (1 - p)] - [\pi \log \pi + (1 - \pi) \log (1 - \pi)]\},$$

and the total expected utility of a particular strategy p is

$$\begin{aligned} U(p) &= U_M + U_I \\ &= a[(1 - \pi) + p(2\pi - 1)] + b[\pi \log p + (1 - \pi) \log (1 - p)] \\ &\qquad - b[\pi \log \pi + (1 - \pi) \log (1 - \pi)]. \end{aligned}$$

The strategy p which maximizes expected utility can be found by setting

$$\frac{dU(p)}{dp} = 0,$$

that is,

$$a(2\pi - 1)p^2 + [b - a(2\pi - 1)]p - b\pi = 0.$$

By dividing by b and letting $\alpha = a/b$,

$$\alpha(2\pi - 1)p^2 + [1 - \alpha(2\pi - 1)]p - \pi = 0. \tag{3a}$$

From (3a) it may be seen that under the conditions of the conventional statistical learning experiment, conditions in which there is no payoff other than the satisfaction of knowing the correctness or incorrectness of a prediction, $\alpha = 0$ and thus $p = \pi$. That is, the Estes prediction is yielded for such conditions. However, if reinforcement is increased, then, setting $\lambda = 2\pi - 1$,

$$p = \frac{-(1 - \alpha\lambda) \pm \sqrt{(1 - \alpha\lambda)^2 + 4\alpha\lambda\pi}}{2\alpha\lambda}. \tag{3b}$$

For three commonly used values of π the predictions that Model III yields concerning a subject's choice and strategy behavior are shown in Figure 1. It is interesting to notice that, in contrast to Models I and II, Model III does not predict that $p = 1$ for any finite value of α; rather, in Model III, p approaches 1 asymptotically as α increases.

Relevant Experimental Evidence

Experimental data are available to support the general approach to the prediction of asymptotes shared by these models as opposed to the general features of statistical association theory. That is, there are data supporting the notion that an increase in the reinforcement contingent upon a correct prediction, and thus the utility of a correct prediction, should result in a subject's p being higher than π, and that under high reinforcement p should approach unity. These data are presented below. They are of interest here for the additional reason that the value of α can be estimated from them and this estimate may be used in future studies to predict p (from Model I).

A study was conducted by Siegel and Goldstein [19] in which the experimental situation already described was used and in which subjects were observed under three conditions of reinforcement: No Payoff, Reward, and Risk. Two lights were illuminated according to a random series, with one illuminating 75 percent of the trials ($\pi = .75$) and the other 25 percent of the trials ($1 - \pi = .25$). The subjects were 36 male students. They were randomly assigned to three sets of equal size, the assignment determining the condition under which the subject would be run. The three conditions were as follows.

No Payoff. Under the No Payoff condition, the reinforcement for each prediction consisted simply of seeing the outcome—seeing whether the right or left bulb lit, and thereby determining whether one's prediction was con-

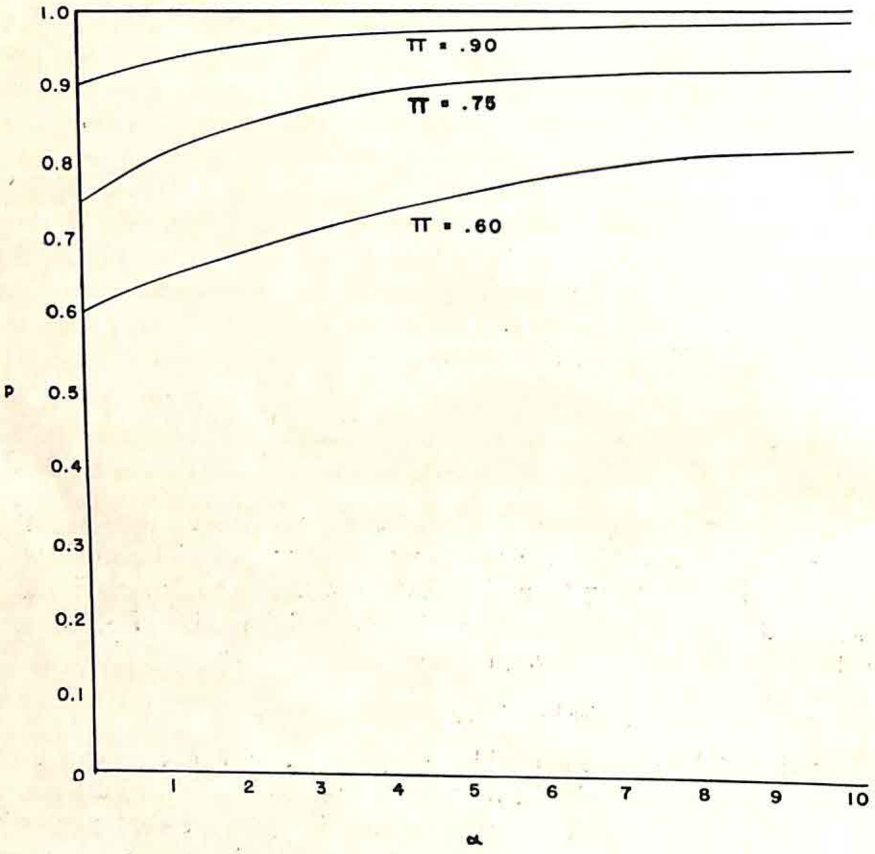


FIGURE 1
Predicted values of p from (3b) with $\alpha = 0$ to 10 for three levels of π.

firmed or disconfirmed. The No Payoff condition has been the "classical" situation for studies of human behavior in two-choice situations.

Reward. Under the Reward condition, the reinforcement for each prediction consisted of seeing the outcome and receiving five cents for each correct prediction. The reward was given at the conclusion of each trial in which S's prediction was confirmed.

Risk. Under the Risk condition, the reinforcement for each prediction consisted of seeing the outcome, receiving five cents if the prediction was confirmed, and losing five cents if it was disconfirmed. At the conclusion of every trial, S either received or forfeited five cents, depending on whether his prediction had been correct or incorrect.

Every subject, regardless of condition, was given 75 cents at the start of the study and told that whatever money he held at the conclusion would be his to keep.

The purpose of the experiment was to test a hypothesis drawn from the decision-making approach introduced earlier, a hypothesis which reflects the general approach of the three models which have been presented. The hypothesis was that *the asymptotic probability of a person's predicting the occurrence of the more frequent event in a two-choice uncertain outcome situation is a function of the level of reinforcement present in the situation, such that the probability of predicting the more frequent event will tend toward unity as the rewards (positive utility) and costs (negative utility) of correct and incorrect predictions are increased.*

That is, using Estes' notation [cf. 6], where $\bar{p}_1(\infty)$ is the mean asymptotic probability of predicting E_1 for a group of like individuals, the hypothesis was that

$$\bar{p}_1(\infty) \text{ under Risk} > \bar{p}_1(\infty) \text{ under Reward} > \bar{p}_1(\infty) \text{ under No Payoff.}$$

The data confirmed the prediction at $p < .001$ [16]. *S*s under the Risk condition predicted the more frequent light oftener than *S*s under the Reward condition, and these in turn predicted the more frequent light oftener than *S*s under the No Payoff condition. Under the No Payoff condition, the mean proportion of times that the more frequent light was predicted during the final 20 trials of the first 100-trial series was .70. Under Reward, the mean was .77, and under Risk it was .93.

For those 12 *S*s (four randomly selected from each group) who were maintained under the same reinforcement conditions for 200 additional trials, the mean proportion of times that the more frequent light was predicted during the final 20 trials of the final 100-trial series was .75 under the No Payoff condition, .86 under the Reward condition, and .95 under the Risk condition. There was no overlap among the scores of these *S*s under the three conditions at the end of 300 trials.

REFERENCES

[1] Bernoulli, D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperiales Petropolitanae*, 1738, 5, 179–192. Translated by L. Sommer, *Econometrica*, 1954, 22, 23–36.

[2] Bush, R. R. and Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

[3] Davidson, D., Siegel, S., and Suppes, P. Some experiments and related theory in the measurement of utility and subjective probability. In D. Davidson, P. Suppes, and S. Siegel, *Decision making: an experimental approach*. Stanford: Stanford Univ. Press, 1957. Pp. 19–81.

[4] Edwards, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417.

[5] Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

[6] Estes, W. K. Individual behavior in uncertain outcome situations: An interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954. Pp. 127–137.

[7] Estes, W. K. and Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

[8] Estes, W. K. and Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

[9] Flood, M. M. Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954. Pp. 287–300.

[10] Galanter, E. and Gerstenhaber, M. On thought: the extrinsic theory. *Psychol. Rev.*, 1956, **63**, 218–228.

[11] Goodnow, J. J. Determinants of choice distributions in two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

[12] Grant, D. A., Hake, H. W., and Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

[13] Hake, H. W. and Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

[14] Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939 **25**, 294–301.

[15] Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

[16] Jonckheere, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

[17] Messick, S. J. and Solley, C. M. Probability learning in children: some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.

[18] Siegel, S. Behavior in a two-choice uncertain outcome situation under the condition of risk. Unpublished study, Stanford Univ., 1953.

[19] Siegel, S. and Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, 57, 37–42.

[20] Simon, H. A. A comparison of game theory and learning theory. *Psychometrika*, 1956, **21**, 267–272.

[21] Toda, M. Guessing sequences under various conditions of payoff. *Jap. psychol. Res.*, 1956, **4**, 11–23.

[22] von Neumann, J. and Morgenstern, O. *Theory of games and economic behavior.* (2nd ed.) Princeton: Princeton Univ. Press, 1947.

Manuscript received 1/30/59
Revised manuscript received 4/14/59

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

THE CONCEPT OF CONFIGURATION OF INTERPERSONAL RELATIONS IN A GROUP AS A TIME-DEPENDENT STOCHASTIC PROCESS

LEO KATZ AND CHARLES H. PROCTOR

MICHIGAN STATE UNIVERSITY

To explain changes in the sociometric configuration of a group through time, a problem arises of the extent to which such changes may be viewed as the aggregation of part-processes occurring at the level of two-person choice structures. A possible model is a Markov chain in which three possible states are mutual choice, one-way choice, and indifference, one realization for each pair of choosing individuals in the group. Choice data for an eighth-grade classroom are fitted to this model and are used to answer questions of constancy of transition probabilities, order of the chain, and sex differences.

While the full complexity of the notion of the configuration of interpersonal choices as representative of a social group has been frequently discussed in the literature, most authors (including the present ones) have quickly abandoned attempts to deal with the process as a whole and have instead concentrated upon particular facets of the process. Although practically everyone recognizes that the process is one which evolves and modifies in time, almost all past investigations have been couched in terms of single points in time. The literature is replete with examples of studies leading to the description of the configuration in a static way; there exist only a very few examples of attempts to deal with as many as two observations of the configuration, separated in time.

In this paper a view of the configuration as a time-dependent process is specified and certain simplified stochastic models are developed, retaining all of the essential features of the process. Finally, an actual example consisting of a series of observations on the configuration for a single group is considered, and tests are applied to see whether the models describe the process reasonably well.

Configuration of Interpersonal Relationships as a Time Process

The standard sociometric definition represents a group in terms of the aggregate of all of the interpersonal relations between pairs of individuals constituting the group. Such an aggregate of relations is called the *configuration* for the group; we recognize that such a description is valid only at a specific point in time. In this sense the configuration of the group is a function of time, and changes with time. As time proceeds the configuration changes kaleidoscopically with some of the interpersonal relations remaining unchanged, while others change abruptly from one quality of the relationship to another.

A particular relation between a designated pair of individuals, at any point in time, is considered to be in one of a discrete set of states representing the various possible qualities of this interpersonal relationship. As time proceeds, this state continues in force until some subsequent point at which it changes abruptly to some other state, continuing in the second state until the next change, which may be a return to the previous state or a passage to a new state.

Alternatively, one may wish to look at the entire configuration consisting of the aggregation of all of the pairwise relations, and consider transitions among the possible states of the configuration. However, if there are N individuals in the group and each relation may be in any of k states, the number of conceivable states of the configuration is $k^{\binom{N}{2}}$, a fantastically large number. With $N = 10$ and $k = 3$, this number is 3×10^{21}.

The Two-Person Relation

Although the pressures on the two-person relationship are continuous from within and without, so changes in the quality of the relation may take place at any point in time, administrative necessity indicates that the group will only be observed at discrete points in time. With this restriction, the only recognized changes will be those in the quality of the relation which may have occurred at any point following the preceding time of observation. Missed completely will be those cases where the relation changes in two steps, perhaps even returning to the previous state. This may not be at all serious since these states are of very brief duration. In what follows these transitory states which fall through the sieve of our observation are ignored.

The possible states in which the relation may be found constitute a double J-nary system in which the two orientations of each asymmetric relation may or may not be identified. The simplest cases consist of double binary relations in which each individual is "on" or "off" with respect to his partner in the pair. It may be desirable to differentiate between the relations "on-off" and "off-on" as between a specified pair of individuals. The differentiation is difficult to interpret, unless there is a complete ordering among the individuals in the group in terms of status, income, etc. In these instances it is essential to know whether a one-way relationship goes from the lower to the higher, or the reverse. If, for the moment, it is assumed that this identification is not desired, then only three states are possible for the two-person relation, the state of mutual positive choice, the state of one-way choice, and the state of non-choice in both directions. For convenience these are designated by the numbers 2, 1, 0, respectively, corresponding to the number of positive choices expressed.

Although it seems almost preposterous to hold the position strongly, logic demands initial consideration of a simple model in which the relations at different periods of time are independent. If this is the case then the probabilities of the relations 0, 1, and 2 are independent of the actual relations observed in earlier periods. When sociometric measurements at two points in time are available for the same group it is possible to test the hypothesis of independence by examining the 3 × 3 array of numbers of pairs falling into the various combinations of relation before by relation afterward. If the numbers observed are sufficiently large, a simple chi-square test with four degrees of freedom is appropriate. Almost certainly, this test will yield rejection of the hypothesis of independence as a result of persistence of large entries in the principal diagonal of the array.

If the process is not random and independent in time, one might next investigate the nature of the dependence of the state of the process at one point in time upon states at other points in time. Although it is possible that a case might be made for a more general mode of dependence, experience in other investigations would indicate that the next most likely model to be tried is some sort of Markov process, with the nature of dependence on the past so far unspecified. A rather complete exposition of the theory of discrete Markov processes may be found in Feller ([2], ch. XV and XVI). At this point it is possible to propose a variety of different tests which vary with the specific nature of the Markov process under consideration. It is not possible to outline all of the tests which might be made; instead, a particular collection of data will be examined and how one may determine the dependence of the process on the past will be indicated by example.

Empirical Examination of the Markov Properties

Data consist of the choices made by an eighth-grade class at four time points in the school year reported by Taba as part of a project in intergroup education [3].* Twenty-five students, 16 girls and 9 boys, were asked, "With whom would you like to sit?" This sociometric test was administered in September and November and in January and May of the following year. All pupils made three choices each except in November when two gave only two choices each. A complete description of the research and an able analysis of the static group structure is available ([3], pp. 45–75).

Frequencies of the various transitions in the 300 two-person relationships between pairs of time points, which appear in Table 1, allow one to test whether there is independence in the sense of uniformity of distribution

---

*The data were used in this study with the kind permission of Dr. Taba.

TABLE I

Distributions of Types of Transitions in Two-Person Choice Structures, Test Population of 25 Eighth Graders

Type of Transition		Time Periods					
Before	After	Sept. to Nov.	Nov. to Jan.	Sept. to Jan.	Jan. to May	Nov. to May	Sept. to May
Mutual to	mutual	7	7	6	6	5	5
	1-way	5	0	3	7	5	3
	indiff.	3	6	6	1	3	7
1-way to	mutual	5	7	6	6	3	3
	same 1-way	17	18	14	15	15	9
	opposite 1-way	2	3	1	1	5	4
	indiff.	21	18	24	25	23	29
Indifference to	mutual	1	0	2	4	8	8
	1-way	22	26	29	20	18	27
	indiff.	217	215	209	215	215	205
Sums		300	300	300	300	300	300

between states of these relations at different times. Tables such as the following were examined (in each case, with 4 degrees of freedom).

State of relation in September		State in November			Sums	
		2	1	0		
Mutual	2	7	5	3	15	
1-way	1	5	19	21	45	$\chi^2 = 121.82$
Indifference	0	1	22	217	240	
Sums		13	46	241	300	

State of relation in November		State in May			Sums	
		2	1	0		
Mutual	2	5	5	3	13	
1-way	1	3	20	23	46	$\chi^2 = 82.34$
Indifference	0	8	18	215	241	
Sums		16	43	241	300	

Without doubt, the later states of two-person relationships are affected by previous states and it remains to be seen whether the dependence is Markovian.

One asks first if the transition probabilities are constant from one pair of time points to the next. For example, do the data contradict the statement that starting from a mutual structure the probability of remaining mutual is the same for the transition September to November as for the

November to January transition? To make the time gap of the transition equal, the transition from September to November is compared with that from November to January and the September to January transition is compared with that from January to May. The test consists of computing a chi-square value, as in a test of association, for the various 2 × 3 and 2 × 4 tables ([1], p. 98). An example is the following.

Starting from a Mutual and ending at . . .	Transition Sept. to Nov.	 Nov. to Jan.	Sums
Mutual	7	7	14
1-way	5	0	5
Indifference	3	6	9
Sums	15	13	28

$\chi^2 = 5.87$

The larger the chi-square value, the more untenable becomes the assumption of constant transition probabilities.

For the September to November and November to January comparison the chi-square values are: starting from mutual, 5.87; from one-way, .80; from indifference, 1.36. The comparison of the September to January with the January to May transition yields: starting from mutual, 5.13; from one-way, .00; from indifference, 2.40. Although there is a suggestion that the probabilities of reaching states from a mutual relation differ from one time to another, the chi-square values do not attain significance at the .05 level. Thus, we will proceed as though they were constant through time.

TABLE 2

Second Order Transitions in Two-Person Choice Structures from September-November to January

Structure in		Structure in January			
September	November	Mutual	1-way	Indifference	Sums
Mutual:	mutual	4	0	3	7
	1-way	2	3	0	5
	indifference	0	0	3	3
1-way:	mutual	2	0	3	5
	1-way	4	9	6	19
	indifference	0	6	15	21
Indifference:	mutual	1	0	0	1
	1-way	1	9	12	22
	indifference	0	20	197	217
Sums		14	47	239	300

We now ask if the Markov chain is of order one or order two. For this test the distinction between same and different one-way was ignored so that there would be only nine transitions between the three states before and after, a more conventional chain. The data used to make the test are found in Tables 2 and 3. The present test likewise consists of making various chi-square tests of association—in this case for 3 × 3 tables. For example, from Table 2 consider the following tabulation.

TABLE 3

Second Order Transitions in Two-Person Choice Structures from September-January to May

Structure in		Structure in May			
September	January	Mutual	1-way	Indifference	Sums
Mutual:	mutual	4	1	1	6
	1-way	1	1	1	3
	indifference	0	1	5	6
1-way:	mutual	2	4	0	6
	1-way	0	5	10	15
	indifference	1	4	19	24
Indifference:	mutual	0	2	0	2
	1-way	5	10	14	29
	indifference	3	15	191	209
Sums		16	43	241	300

Structure in		Structure in January			
Sept.	Nov.	Mutual	1-way	Indifference	Sums
Mutual	Mutual	4	0	3	7
1-way	Mutual	2	0	3	5
Indifference	Mutual	1	0	0	1
	Sums	7	0	6	13

$\chi^2 = 1.26$ (2 degrees of freedom because of 0's)

A large value of chi-square would indicate that a difference in the state of the structures in September has influence over the movement to new structures in January. That is, larger chi-square values indicate a second or higher order chain. The observed chi-square values are summarized in Table 4.

Obviously the chain of four-month gaps, September to January to May, does not show second-order dependence. However, for the chain of two-month gaps, the state of the structure in September appears to exert influence over the way in which the indifferences of November changed into

TABLE 4

Values of Chi-Square for Tests of Order of Chain

Structure at Intermediate Time Point	September-November to January Transition		September-January to May Transition	
	Chi-Square	Degrees of Freedom	Chi-Square	Degrees of Freedom
Mutual	1.26	2	5.78	4
1-way	8.02	4	4.26	4
Indifference	7.82*	2	4.27	4
Sums	17.25*	8	14.31	12

*Significant at 5% level. The value 7.82 when corrected for continuity becomes 6.83, which is still significant at the 5% level. This correction consists of computing the next smallest value of chi-squared, in this case 5.83, for the same set of marginal totals and using as "corrected" value the mid-point of the two, or 6.83.

other states in January. The following tabulation shows the direction of this influence (numbers in brackets are theoretical or expected frequencies).

Structure in		Structure in January			
September	November	Mutual	1-way	Indifference	Sums
Mutual	Indifference	0	0 [0.32]	3 [2.67]	3
1-way	Indifference	0	6 [2.26]	15 [18.73]	21
Indifference	Indifference	0	20 [23.41]	197 [193.59]	217
	Sums	0	26	215	241

Apparently, the passage through indifference does not eliminate the tendency to return to the previous state. A similar but less marked tendency can be noted in the chain of four-month gaps. Possibly the investigation of more closely spaced time points would reveal a considerable number of returns to a previous state even upon passage through a mutual or a one-way structure.

Having decided that the chain of four-month gaps is of order one and that the transition probabilities are stable, the data in Table 1 are used to estimate these probabilities. From a mutual structure the probabilities of becoming in four months a mutual, a one-way, or an indifference are .414, .345, and .241, respectively; from a one-way to a mutual, same one-way, different one-way, or indifference are .130, .315, .022, and .533, while from an indifference to a mutual, one-way, or indifference are .013, .102, and .885, respectively. The matrix of transition probabilities is constructed by combining the "to same" and "to different one-way" transitions.

From		To 2	To 1	To 0
Mutual	2	.414	.345	.241
1-way	1	.130	.337	.533
Indifference	0	.013	.102	.885

Since all states are accessible from each state, the Markov chain is irreducible and possesses a unique stationary distribution with limiting probabilities for mutual, one-way, and indifference of .051, .149, and .800, respectively. Among the 300 two-person structures in a 25-person group, one would expect to find as "stable" values 15 mutuals, 45 one-ways, and 240 indifferences. It is thus not surprising that the numbers of mutuals, one-ways, and indifferences were observed to be so stable. The numbers of mutuals were 15 in September, 13 in November, 14 in January, and 16 in May; the one-ways were 45, 46, 47, and 43; while the indifferences were 240, 241, 239, and 241. What was perhaps surprising was the fact that the "initial" or September distribution (15–45–240) conformed exactly to the limiting values.

Girl and Boy Subgroups

Taba [3] notes a sex cleavage and consequently the data on transitions were examined separately for boys and girls. The numbers of transitions

TABLE 5

Distributions of Types of Transitions in Two-Person Choice Structures, Boy to Boy, Girl to Girl and Cross Sex by Five Pairs of Time Points

Type of Transition	Sept. to Nov.			Nov. to Jan.			Sept. to Jan.			Jan. to May			Sept. to May		
	B to B	G to G	X Sex	B to B	G to G	X Sex	B to B	G to G	X Sex	B to B	G to G	X Sex	B to B	G to G	X Sex
Mutual to...															
Mutual	2	5	0	3	4	0	1	5	0	2	4	0	2	3	0
1-way	3	2	0	0	0	0	3	0	0	1	5	1	2	1	0
Indiff.	1	2	0	3	3	0	2	4	0	0	1	0	2	5	0
1-way to...															
Mutual	4	1	0	0	6	1	2	3	1	1	3	2	2	1	0
Same	2	15	0	8	9	1	2	10	2	3	7	5	2	6	1
Opposite	1	1	0	1	2	0	0	0	1	0	1	0	1	3	0
Indiff.	3	11	7	4	13	1	6	15	3	6	11	8	5	18	6
Indiff. to...															
Mutual	0	1	0	0	0	0	0	2	0	2	1	1	1	4	3
1-way	7	12	3	1	11	14	5	12	12	5	11	4	4	14	9
Indiff.	13	70	134	16	72	127	15	69	125	16	76	123	15	65	125
Sums	36	120	144	36	120	144	36	120	144	36	120	144	36	120	144

were counted for the 36 pair-relations among the 9 boys and for the 120 relations among the 25 girls. These appear in Table 5. Due to the fact that more than 85 percent of the 144 cross-sex pairs were consistently indifferences, the choice behavior of the whole class can almost be considered as the sum of two disjoint subgroups.

Tests for constancy of transition probabilities from a two-month or four-month gap to the next and for second- versus first-order dependency were performed for the girl-to-girl choices. No departure from constancy nor from a first-order chain were detected; however, the number of cases is small. Tables 6 and 7, which parallel Tables 2 and 3 above, are given for completeness.

TABLE 6

Second Order Transitions in Girls' Two-Person Choice Structures from September-November to January

Structure in		Structure in January			Sums
September	November	Mutual	1-way	Indifference	
Mutual:	mutual	3	0	2	5
	1-way	2	0	0	2
	indifference	0	0	2	2
Mutual:	mutual	0	0	1	1
	1-way	3	7	6	16
	indifference	0	3	8	11
Mutual:	mutual	1	0	0	1
	1-way	1	4	7	12
	indifference	0	8	62	70
Sums		10	22	87	120

TABLE 7

Second Order Transitions in Girls' Two-Person Choice Structures from September-January to May

Structure in		Structure in May			Sums
September	January	Mutual	1-way	Indifference	
Mutual:	mutual	3	1	1	5
	1-way	0	0	0	0
	indifference	0	0	4	4
1-way:	mutual	1	2	0	3
	1-way	0	4	6	10
	indifference	0	3	12	15
Indifference:	mutual	0	2	0	2
	1-way	3	4	5	12
	indifference	1	8	60	69
Sums		8	24	88	120

Of interest is the similarity of the boys and girls with respect to the relative frequencies of the transitions. To show this, tables such as the following were examined for the combined (September to November with November to January) two-month transitions.

Starting from a Mutual and ending two months later at . . .	Sub-group			
	Boys	Girls	Sums	
Mutual	5	9	14	
1-way	3	2	5	$\chi^2 = .90$
Indifference	4	5	9	
Sums	12	16	28	

For the two-month gap the chi-squares are: starting from mutual, .90; starting from one-way, 1.21; starting from indifference, 1.60. For the four-month gap they were: starting from mutual, .95; starting from one-way, 1.04; and starting from indifference, 4.06. These values show a strong degree of similarity between the two subgroups. Although the number of relations within the boys' subgroup is only 30 percent of that within the girls' group while the boys made 56 percent as many choices as did the girls' subgroup, the boys initiated over half of the intersex choices, thus equalizing the ratios of number of in-group choices to number of possible in-group relations for the two subgroups. Over and above this point of similarity, the relative amounts of exploratory and persisting choice behaviors are similar for the two subgroups.

Summary

After proposing that interpersonal choice behavior viewed through time may conform to a Markov chain model, some data from an eighth-grade class choosing on a seating criterion were examined. Within the limits of the numbers of cases available no evidence was found to suggest that the transition probabilities varied with time. For the two-month gaps there may possibly be second-order dependence of the chain, but the chain of four-month gaps did not depart significantly from first-order dependence. Interestingly, the initial or September numbers of mutuals, one-ways, and indifferences conformed exactly to the limiting distribution for a first-order chain of four-month gaps.

The girls' and boys' subgroups behaved similarly with respect to the amounts of persisting or repetitive choosing and of exploratory choosing. No departures from the first-order Markov chain model achieved statistical significance among the within-sex choices.

REFERENCES

[1] Anderson, T. W. and Goodman, L. A. Statistical inference about Markov chains. *Ann. math. Statist.*, 1957, 28, 89–110.

[2] Feller, W. *An introduction to probability theory and its applications*. Vol. 1. (2nd ed.) New York: Wiley, 1957.

[3] Taba, H. With perspective on human relations: a study of peer group dynamics in an eighth grade. Washington, D. C.: Amer. Council Educ., 1955.

Manuscript received 1/24/59

Revised manuscript received 5/16/59

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

PATTERN VARIANTS ON A SQUARE FIELD

S. J. PROKHOVNIK*

THE UNIVERSITY OF NEW SOUTH WALES

A quantitative approach to the psychology of pattern recognition requires knowledge of the number of possible variants of any particular pattern. The general solution for the number $[p/m^2]$ of pattern variants that p counters can form on a square network of m^2 positions is obtained by elementary group theory. The exact solution is given in terms of the different types (symmetric, asymmetric, etc.) of patterns possible and an approximate formula for the total number of patterns is also developed.

Given a number p of identical counters and a square network of m^2 positions to place them on, it is possible to form a finite number of configurations given by $\binom{m^2}{p}$. However, if only different patterns are allowed, the number is reduced considerably, since certain sets of configurations differ from one another only by rotation and/or reflection of the field; each such set represents a single pattern. The number of elements in the set will vary according to the type of symmetry or asymmetry of the particular pattern. Solution for the number of pattern variants requires differentiation among four types of configurations as well as a number of subtypes.

For simple cases, particular solutions can be obtained, though laboriously, by means of elementary arithmetic and geometry. Application of group theory enables a general solution applicable in all cases. The solution is extended with comparative ease to the case where the pattern consists not of identical counters but of two or more different types. This requires only the application of a permutation factor which will generally differ for each of the four types of configurations. The problem has found (in fact arose from) practical application in certain psychological experiments, work by S. Kamocki, not yet published. It may also be useful, using three-dimensional networks, in physico-chemical problems involving patterns of discrete particles, for example the pattern of amino acids in the protein molecule or the configurations of other molecules or physical structures.

Problem in Terms of Postulates

The number of ways of placing p identical counters on a square network of m^2 positions is defined to be $[p/m^2]$, where configurations which differ from one another by a rotation and/or reflection are considered identical.

---

*The author is indebted to Mr. J. Sandiford for his invaluable assistance on the theoretical aspects of the problem and to Mr. J. L. Griffith for his helpful criticism.

We may establish a coordinate system on the network, with the axes dividing the squares into four quadrants. Taking $m = 2n$ for even networks and $m = 2n + 1$ for odd networks $(n = 1, 2, 3, \cdots)$, then the range of x, y will be

$$\left(-\frac{m-1}{2},\ \frac{m-1}{2}\right).$$

Define the operation R as a counterclockwise rotation through 90°. Thus $R(x, y) = (-y, +x)$, $R^4 = I$. Define the operation S as a reflection in the middle vertical line, i.e., the y-axis. Thus $S(x, y) = (-x, y)$, $S^2 = I$. The number of systems which are not equivalent under S, R or any combination of S and R is required.

Properties of the Group $\{R, S\}$

Denote $\{R, S\}$, a group whose elements are R, S and the identity element I, as the group G_8 which is of order 8. Thus $G_8 = \{R, S\} = I, R, R^2, R^3; S, SR, SR^2, SR^3$, where by convention the double operation SR represents the operation R followed by the operation S.

Note that R, S are not commutative but that $SR = R^3S$ is a reflection on the diagonal of positive gradient; $SR^3 = RS$ is a reflection on the other diagonal; and $SR^2 = R^2S$ is a reflection on the horizontal axis. Hence the Group G_8 is complete as defined above. G_8 has three subgroups of order 4, viz., the cyclic group

$$\begin{aligned} G_{41} &= \{R\} &&= I, R, R^2, R^3, \\ G_{42} &= \{R^2, S\} &&= I, S, SR^2, R^2, \\ G_{43} &= \{R^2, SR\} &&= I, SR, R^2, SR^3; \end{aligned}$$

and the five cyclic subgroups of order 2 denoted by

$$G_{21} = \{R^2\},\ G_{22} = \{S\},\ G_{23} = \{SR^2\},\ G_{24} = \{SR\},\ G_{25} = \{SR^3\}.$$

Types of Configurations Possible

Define symmetric, semi-symmetric, anti-symmetric, and asymmetric configurations in terms of their group properties. A pattern may then be defined in terms of the various configurations as follows: if C is a configuration, then all the elements of g_8C form a single pattern, where g_8 represents any operation in the group G_8. The number of p-position patterns in a network of m^2 positions has been defined as $[p/m^2]$, which is to be found. Proofs of the theorems mentioned below are based on elementary group theory; the longer of these appear in the Appendix, referred to by (A).

Symmetric configurations are those invariant under every combination of R, S, i.e., under the 8-order group G_8. This type of configuration is demonstrated by Figure 1 and is unique, each one representing also a unique pattern. $[G_8]$ will designate the number of patterns of this type in any particular case.

Semi-symmetric configurations (Figures 2, 3, and 4) are those invariant under the subgroups of order 4, but not under G_8. It is easily shown (A) that there are two, and only two, configurations of this type, (each invariant under the same subgroup of order 4) based on the same pattern, i.e., a semi-symmetric pattern. Thus the number of semi-symmetric patterns is $[G_4]$ and the number of semi-symmetric configurations is $2[G_4]$. It will be found convenient to denote the number of patterns in the three subgroups of order 4 by $[G_{41}]$, $[G_{42}]$, $[G_{43}]$ pertaining to pattern types exemplified by Figures 2, 3, and 4, respectively.

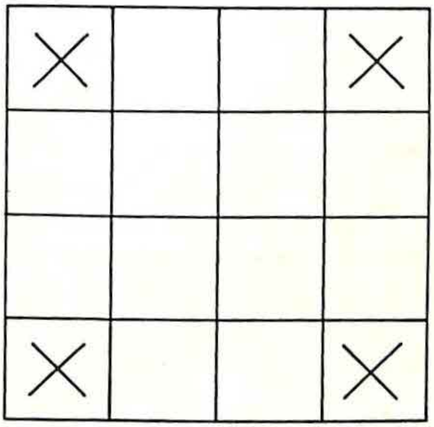


FIGURE 1

A Symmetric G_8 Pattern

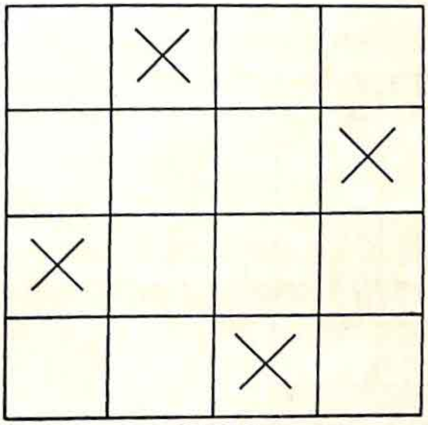


FIGURE 2

A Semi-symmetric G_{41} Pattern

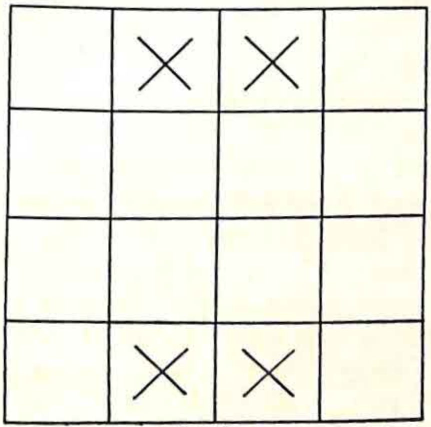


FIGURE 3

A Semi-symmetric G_{42} Pattern

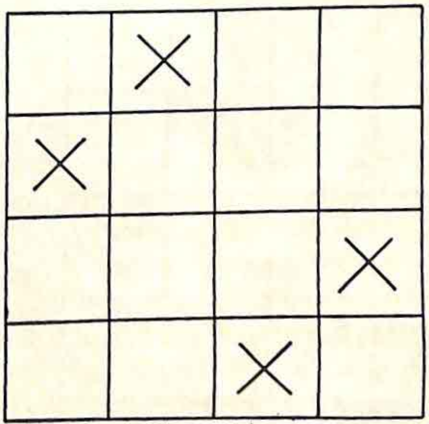


FIGURE 4

A Semi-symmetric G_{43} Pattern

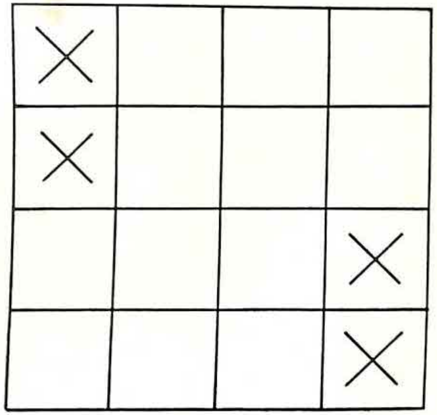


FIGURE 5

An Anti-symmetric G_{21} Pattern

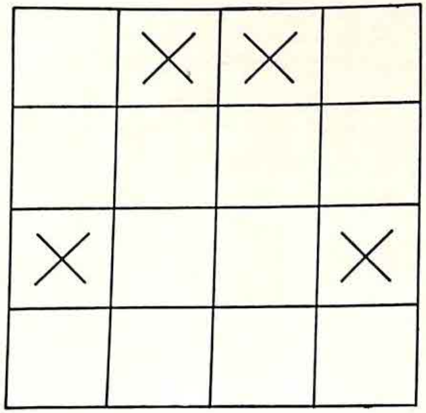


FIGURE 6

An Anti-symmetric G_{22} Pattern

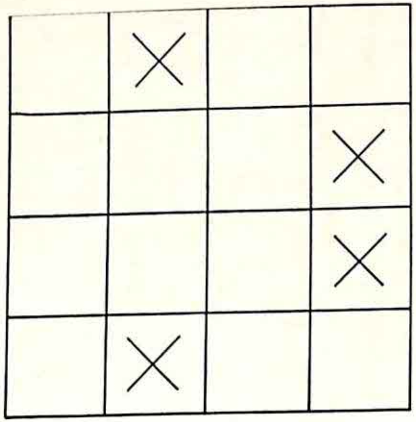


FIGURE 7

An Anti-symmetric G_{23} Pattern

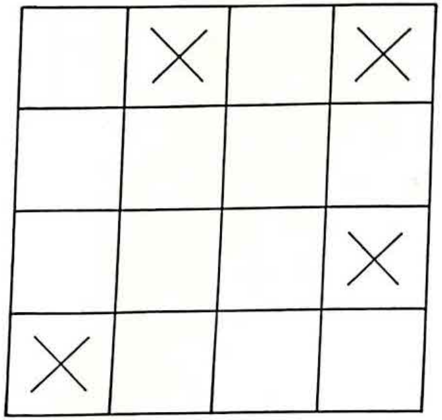


FIGURE 8

An Anti-symmetric G_{24} Pattern

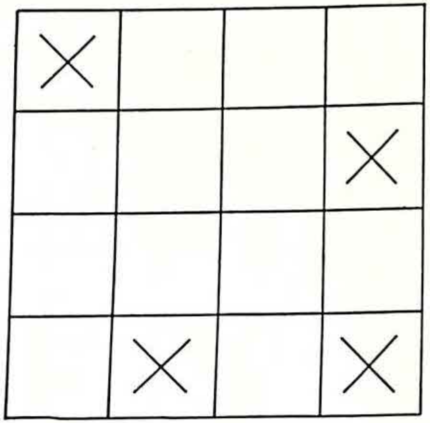


FIGURE 9

An Anti-symmetric G_{25} Pattern

Anti-symmetric configurations (Figures 5, 6, 7, 8, and 9) are those invariant under the subgroups of order 2, but not under any previous groups. There will be four, and only four, such configurations based on the same pattern, i.e., an anti-symmetric pattern (A). Thus the number of anti-symmetric patterns is $[G_2]$, and the number of anti-symmetric configurations is $4[G_2]$. The number of patterns due to the five subgroups of order 2 will be denoted by $[G_{21}]$, $[G_{22}]$, $[G_{23}]$, $[G_{24}]$, $[G_{25}]$ pertaining to the pattern types exemplified by Figures 5, 6, 7, 8, and 9, respectively.

Asymmetric configurations (e.g., Figure 10) can be considered those invariant only under I, or, alternatively, those which do not belong to the 2-, 4-, or 8-order groups of G. Every operation of the group G will yield a different configuration; there are eight configurations for each asymmetric pattern. The number of such patterns $[G_1]$ for the case $[p/m^2]$ will be given by

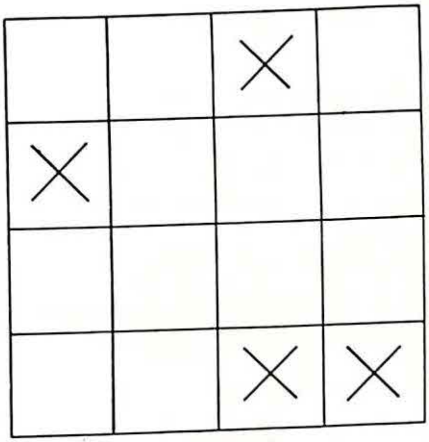


FIGURE 10

An Asymmetric G_1 Pattern

$$[G_1] = \frac{\binom{m^2}{p} - [G_8] - 2[G_4] - 4[G_2]}{8}.$$

Preliminary Theorems

The following theorems, which follow directly from the postulates above, are useful in the solution of the problem.

THEOREM I.

$$\left[\frac{m^2 - p}{m^2}\right] = \left[\frac{p}{m^2}\right],$$

since for every pattern of p counters on an m^2 network there is a corresponding pattern of $(m^2 - p)$ blanks and vice versa.

THEOREM II. *For every configuration invariant under $G_{22} = \{S\}$ there is an equivalent one invariant under $G_{23} = \{SR^2\}$ and vice versa* (A). Figures 6 and 7 exemplify a pair of such configurations.

Hence

$$[G_{22}] = [G_{23}] = [G_{22} + G_{23}].$$

From similar considerations, (e.g., Figures 8 and 9)

$$[G_{24}] = [G_{25}] = [G_{24} + G_{25}].$$

THEOREM III. *In any odd network the number of patterns for $[G_8]$, $[G_4]$, and $[G_{21}]$ will be the same for $p = 2r + 1$ (counters) as it is for $p = 2r$* (A).

Method

It is clear that different considerations apply for odd and even cases, both for networks and for number of points forming the pattern. Further, symmetrical (G_8) patterns of p points over an m^2 network are possible only if p is a multiple of 4 and, in the case when m is odd, also if $p - 1$ is a multiple of 4. Hence solutions of $[p/m^2]$ are given for $m = 2n, 2n + 1$, and $p = 4r, 4r + 1$, where $r, n = 0, 1, 2, \cdots$.

Now any part of a solution for $4r$ or $4r + 1$ counters which involves only multiples of $2r$ is applicable for all even ($p = 2s$) or all odd ($p = 2s + 1$) cases respectively, on the same network. This leads to the following theorem.

THEOREM IV. *If we substitute $2r = s$ in the solutions of $[4r/m^2]$ which involve only multiples of $2r$, then we obtain the solutions of $[2s/m^2]$ when $s = 1, 3, 5, \cdots$. A similar relationship applies between $[(4r + 1)/m^2]$ and $[(2s + 1)/m^2]$.*

PROOF. It is clear that the relationship is necessary. It is also sufficient since if $s = 2r$ in the solutions of $[2s/m^2]$ one obtains corresponding solutions of $[4r/m^2]$. Hence all the solutions of $[2s/m^2]$ are derived according to the theorem from those of $[4r/m^2]$.

Following Theorem IV, the solutions for $p = 2s, 2s + 1$ ($s = 1, 3, 5, \cdots$) are given by replacing $2r$ by s in the starred (*) set of the relevant solutions for $p = 4r, 4r + 1$, respectively. The starred sets are then the complete solutions for the cases where p or $p - 1$ is even but not a multiple of 4.

The calculation can be simplified by noting that for odd networks ($m = 2n + 1$) the argument with respect to diagonals is identical to that with respect to axes. Hence, in such networks,

$$[G_{42}] = [G_{43}]; \qquad [G_{22} + G_{23}] = [G_{24} + G_{25}].$$

In deriving the general solutions, a set of m points invariant under a particular group or subgroup of G can be considered as being composed of subsets of single points, pairs, quartets, and octets invariant under the same group or subgroup. The solution then consists of all combinations of such subsets invariant under the particular group considered. For instance 8-point invariant sets occur only in the symmetric (G_8) patterns, being quartets of pairs of points symmetric about a diagonal or axis. Thus the solution for $[G_8]$ consists of (all possible combinations of) quartets of diagonal or axial points and/or the symmetric octets as described above.

Numerical evaluations of the general solutions for a number of patterns and networks are given in Table 1. The results agree with arithmetical enumeration methods which are possible though laborious for small n, r.

General Solutions

(i) Solution for $[4r/(2n)^2]$.

$$[G_8] = \binom{n}{r} + \binom{n}{r-2}\binom{\frac{1}{2}(n^2-n)}{1} + \binom{n}{r-4}\binom{\frac{1}{2}(n^2-n)}{2} + \cdots \text{ to } \binom{n}{0}\binom{\frac{1}{2}(n^2-n)}{\frac{1}{2}r} \text{ or } \binom{n}{1}\binom{\frac{1}{2}(n^2-n)}{\frac{1}{2}(r-1)},$$

that is,

$$\binom{\frac{1}{2}(n^2-n)}{\frac{1}{2}r} \text{ or } n\binom{\frac{1}{2}(n^2-n)}{\frac{1}{2}(r-1)}.$$

$$[G_4] \ : [G_{41}] = \tfrac{1}{2}\left\{\binom{n^2}{r} - [G_8]\right\},$$

$$[G_{42}] = \tfrac{1}{2}\left\{\binom{n^2}{r} - [G_8]\right\},$$

$$[G_{43}]^* = \tfrac{1}{2}\left\{\binom{2n}{2r} + \binom{2n}{2r-2}\binom{n^2-n}{1} + \binom{2n}{2r-4}\binom{n^2-n}{2} + \cdots - [G_8]\right\}.$$

$$[G_2]^* : [G_{21}] = \tfrac{1}{4}\left\{\binom{2n^2}{2r} - [G_8] - 2[G_4]\right\},$$

$$[G_{22} + G_{23}] = \tfrac{1}{2}\left\{\binom{2n^2}{2r} - [G_8] - 2[G_{42}]\right\},$$

$$[G_{24} + G_{25}] = \tfrac{1}{2}\left\{\binom{2n^2-n}{2r} + \binom{2n}{2}\binom{2n^2-n}{2r-1} + \binom{2n}{4}\binom{2n^2-n}{2r-2} + \cdots - [G_8] - 2[G_{43}]\right\}.$$

$$[G_1]^* = \tfrac{1}{8}\left\{\binom{4n^2}{4r} - [G_8] - 2[G_4] - 4[G_2]\right\}.$$

(ii) Solution for $\left[\frac{4r+1}{(2n)^2}\right]$.

$[G_8]^*$: Nil.

$[G_4]^*$: Nil.

$[G_2]^*$: $[G_{21}]$, $[G_{22} + G_{23}]$, Nil.

$$[G_{24} + G_{25}] = \tfrac{1}{2}\left\{\binom{2n}{1}\binom{2n^2 - n}{2r} + \binom{2n}{3}\binom{2n^2 - n}{2r - 1} + \binom{2n}{5}\binom{2n^2 - n}{2r - 2} + \cdots\right\}.$$

$$[G_1]^* = \tfrac{1}{8}\left\{\binom{4n^2}{4r + 1} - 4[G_2]\right\}.$$

(iii) Solution for $\left[\frac{4r}{(2n+1)^2}\right]$.

$$[G_8] = \binom{2n}{r} + \binom{2n}{r - 2}\binom{\frac{1}{2}(n^2 - n)}{1} + \binom{2n}{r - 4}\binom{\frac{1}{2}(n^2 - n)}{2} + \cdots .$$

$$[G_4] : [G_{41}] = \tfrac{1}{2}\left\{\binom{n^2 + n}{r} - [G_8]\right\},$$

$$[G_{42}]^* + [G_{43}]^* = \binom{2n}{2r} + \binom{2n}{2r - 2}\binom{n^2}{1} + \binom{2n}{2r - 4}\binom{n^2}{2} + \cdots - [G_8].$$

$$[G_2]^* : [G_{21}] = \tfrac{1}{4}\left\{\binom{2n^2 + 2n}{2r} - [G_8] - 2[G_4]\right\},$$

$$[G_{22} + G_{23}] + [G_{24} + G_{25}] = \binom{2n^2 + n}{2r} + \binom{2n + 1}{2}\binom{2n^2 + n}{2r - 1} + \binom{2n + 2}{4}\binom{2n^2 + n}{2r - 2} + \cdots - [G_8] - 2[G_{42}].$$

$$[G_1]^* = \tfrac{1}{8}\left\{\binom{(2n + 1)^2}{4r} - [G_8] - 2[G_4] - 4[G_2]\right\}.$$

(iv) Solution for $\left[\frac{4r+1}{(2n+1)^2}\right]$.

$$\left.\begin{array}{ll}[G_8] & \\ [G_4] & : [G_{41}], \\ & \quad [G_{42}]^* + [G_{43}]^*. \\ [G_2]^* & : [G_{21}],\end{array}\right\} \text{ as for (iii)—Theorem III.}$$

$$\begin{aligned}&[G_{22} + G_{23}] + [G_{24} + G_{25}] \\ &\qquad = \binom{2n+1}{1}\binom{2n^2+n}{2r} + \binom{2n+1}{3}\binom{2n^2+n}{2r-1} \\ &\qquad + \binom{2n+1}{5}\binom{2n^2+n}{2r-2} + \cdots - [G_8] - 2[G_{42}].\end{aligned}$$

$$[G_1]^* = \tfrac{1}{8}\left\{\binom{(2n+1)^2}{4r+1} - [G_8] - 2[G_4] - 4[G_2]\right\}.$$

Ratio of Symmetric to Total Number of Patterns

Table 1 gives a selection of numerical evaluations of $[G_8], [G_4], [G_2], [G_1]$, and the total number of patterns for networks up to 6^2 positions. The total number of configurations $\binom{m^2}{p}$ for each case is also listed. It is clear from Table 1 that the total number of symmetrical patterns $[G_8] + [G_4] + [G_2]$ is only a small ratio of all patterns possible for a particular case, and further that this ratio decreases with increasing m, p according to a square root law.

The latter also follows more rigorously by examining the nature of the particular and general solutions outlined above, whence it is seen that $\{[G_2] + [G_4] + [G_2]\}$ is, in the main, a function of $\binom{\frac{1}{2}m^2}{\frac{1}{2}p}$ and hence of m^p and lower powers of m, while $\binom{m^2}{p}$ is a function of m^{2p} and lower powers. Further it is also easily shown that

$$\binom{\frac{1}{2}m^2}{\frac{1}{2}p} \leq \sqrt{\binom{m^2}{p}} \leq \sqrt{p}\binom{\frac{1}{2}m^2}{\frac{1}{2}p}.$$

Thus it would appear, both from the trend apparent in Table 1 and from the nature of the solution formulas, that $\{[G_8] + [G_4] + [G_2]\}$ approximates $\sqrt{\binom{m^2}{p}}$ rather more for cases where the symmetry is more pronounced (as for odd networks in general) and rather less for the other cases (as for odd positions in even networks, where, for example, there is no contribution from $[G_8]$, $[G_4]$, $[G_{21}]$, and $[G_{22} + G_{23}]$).

The small ratio of symmetric to total number of patterns leads to two generalizations of practical consequence. First, in any physical problem in-

TABLE 1

Numerical Evaluations for Small m , p

m	p	$[G_8]$	$[G_4]$	$[G_2]$	$[G_1]$	Total number of patterns	Total number of configurations $\binom{m^2}{p}$
1	1	1	0	0	0	1	1
2	1	0	0	1	0	1	4
2	2	0	1	1	0	2	6
3	1	1	0	2	0	3	9
3	2	0	2	4	2	8	36
3	3	0	2	8	6	16	84
3	4	2	0	11	10	23	126
4	1	0	0	2	1	3	16
4	2	0	2	9	10	21	120
4	3	0	0	14	63	77	560
4	4	2	5	38	207	252	1,820
4	5	0	0	42	525	567	4,368
4	6	0	6	91	954	1,051	8,008
4	7	0	0	70	1,395	1,465	11,440
4	8	2	10	112	1,550	1,674	12,870
5	1	1	0	4	1	6	25
5	2	0	4	17	28	49	300
5	3	0	4	57	258	319	2,300
5	4	4	7	152	1,503	1,666	12,650
5	5	4	7	328	6,475	6,814	53,130
5	6	0	20	645	21,810	22,475	177,100
5	7	0	20	1,085	59,540	60,645	480,700
5	8	7	28	1,712	134,333	136,080	1,081,575
5	9	7	28	2,372	254,178	256,585	2,042,975
6	1	0	0	3	3	6	36
6	2	0	3	24	66	93	630
6	3	0	0	55	865	920	7,140
6	4	3	15	264	7,227	7,509	58,905
6	5	0	0	468	46,890	47,358	376,992
6	6	0	28	1,698	242,618	244,344	1,947,792
6	7	0	0	2,460	1,042,230	1,044,690	8,347,680
6	8	6	87	7,062	3,778,989	3,786,144	30,260,340
6	9	0	0	8,960	11,763,430	11,772,390	94×10^6
6	10	0	108	21,468	31,762,596	31,784,172	254×10^6
6	11	0	0	24,024	75,088,650	75,112,674	601×10^6

volving the elucidation of the pattern of a structure, it would be clearly advantageous to seek any evidence of symmetry within the structure. If symmetry is displayed, not only is the number of patterns possible reduced significantly, but also these can be listed and checked systematically under the groups and subgroups of G_8, G_4, and G_2. Second, for $p, m > 2$ the total number of patterns $[p/m^2]$ is of the order of $\frac{1}{8}\binom{m^2}{p}$ though always greater than this number. Hence as a second approximation

$$\left[\frac{p}{m^2}\right] = \tfrac{1}{8}\binom{m^2}{p}(1+\epsilon),$$

where ϵ is a small fraction and $\epsilon \to 0$ as $m, p \to \infty$.

It may be shown that a good approximation for ϵ is given by

$$4 \Big/ \sqrt{\binom{m^2}{p}}\,;$$

for if

$$\left[\frac{p}{m^2}\right] = [G_1] + [G_2] + [G_4] + [G_8],$$

where $[G_4]$, $[G_8]$ are very small compared to $[G_2]$ (especially for large m, p), then

$$\left[\frac{p}{m^2}\right] \approx \frac{\binom{m^2}{p} - 4\sqrt{\binom{m^2}{p}}}{8} + \sqrt{\binom{m^2}{p}}$$

$$= \tfrac{1}{8}\binom{m^2}{p}\left(1 + 4 \Big/ \sqrt{\binom{m^2}{p}}\right).$$

This formula gives a good approximation for the total number of patterns as may be seen from Table 2, but it is, of course, no substitute for the general solution when an exact enumeration of the different types of patterns is required.

APPENDIX

Proof that every semi-symmetric pattern has two and only two configurations

Consider a configuration C invariant under G_{41} but not under G_8, i.e.,

$$C = RC = R^2C = R^3C.$$

Now consider the configuration

$$F = SC;$$

then

$$RF = RSC = SR^3C = SC = F,$$

$$R^2F = R(RF) = F,$$

and

$$R^3F = R(R^2F) = F,$$

TABLE 2

Error of Approximation Formula for Small m , p *

m	p	ε (actual)	ε (estimated)	Percentage error in (1 + ε)
3	2	7/9	6/9	-6
3	3	11/21	9/21	-6
3	4	6/13	7/13	5
4	2	2/5	4/11	-3
4	3	1/10	1/6	6
4	4	3/28	3/32	-2
4	5	1/26	2/33	2
4	6	1/20	1/22	- .5
4	7	1/41	1/27	1.3
4	8	1/25	1/28	- .4
5	2	1/3	3/13	-8
5	3	1/9	1/12	-2.5
5	4	1/19	1/28	-1.6
5	5	1/38	1/58	- .9
5	6	1/66	1/105	- .5
5	7	1/108	1/173	- .3
5	8	1/153	1/260	- .3
5	9	1/212	1/357	- .2
6	2	2/11	4/25	-2
6	3	1/32	1/21	2
6	4	1/50	1/61	- .4
6	5	1/199	1/153	.2
6	6	1/280	1/349	- .08
6	7	1/848	1/722	.02
6	8	1/1050	1/1375	- .02
6	9	1/2627	1/2425	.002
6	10	1/2940	1/4000	- .009
6	11	1/6250	1/6125	.0001

*ϵ is estimated by $4 \Big/ \sqrt{\binom{m^2}{p}}$. The actual value of ϵ is determined from

$$\left[\frac{p}{m^2}\right] = \frac{1}{8}\binom{m^2}{p}(1+\epsilon),$$

where $[p/m^2]$ is known. Both actual and estimated values are presented in round figures, where accuracy is not severely affected. The final column gives percentage error in $(1+\epsilon)$ and thus in $[p/m^2]$ by use of ϵ (estimated).

therefore

$$F = RF = R^2F = R^3F.$$

If also $F = SC = C$, then $C = SRC = SR^2C = SR^3C$, i.e., both C and F would be identical configurations invariant under G_8 and hence previously included. If however $F = SC \neq C$, then F and C are different configurations based on the same pattern, and both are invariant under G_{41} but not under G_8 . Also the invariance of C and F involves all the elements of g_8C , therefore there are no other configurations based on this pattern. Similar proofs apply to configurations invariant under G_{42} , G_{43} .

Proof that every anti-symmetric pattern has four and only four configurations

Consider a configuration C_1 invariant under G_{21} but not under G_8 , G_{41} , G_{42} , G_{43} , i.e., $C_1 = R^2C_1$. Now consider the configuration $C_2 = RC_1$;

$$R^2C_2 = R^3C_1 = R(R^2C_1) = RC_1 = C_2 .$$

If also $C_2 = RC_1 = C_1$, then $C_1 = RC_1 = R^2C_1 = R^3C_1$, i.e., both C_1 , C_2 would be identical configurations invariant under G_{41} and hence previously included.

If, however, $C_2 = RC \neq C_1$ then C_1 , C_2 are different configurations, based on the same pattern and both invariant under G_{21} but not under any other group of higher order. It may be shown similarly that there exist two other configurations, viz.

$$C_3 = SRC_1 \quad \text{and} \quad C_4 = SC_1 ,$$

invariant under G_{21} but under no other group of higher order.

Hence C_1 , C_2 , C_3 , C_4 are four different configurations, all invariant under G_{21} and based on the same pattern. Also the invariance of C_1 , C_2 , C_3 , C_4 involves all the elements of g_8C_1 ; therefore there can be no other configurations based on this pattern. Similar proofs apply to configurations invariant under G_{22} , G_{23} , G_{24} , G_{25} .

Proof of Theorem II. Consider a configuration C_1 invariant under G_{22} , i.e., $SC_1 = C_1$; then there will always exist a different configuration $C_2 = RC_1$; but $SR^2C_2 = SR^3C_1 = RSC_1 = RC_1 = C_2$, therefore C_2 is invariant under $G_{23} = \{SR^2\}$. Consider now a third configuration $C_3 = RC_2 = R^2C_1$; then $SC_3 = SR^2C_1 = R^2SC_1 = R^2C_1 = C_3$, therefore C_3 is also invariant under G_{22} . Similarly it may be shown that $C_4 = R^2C_1$ is also invariant under G_{23} . These four configurations and their invariants exhaust all the elements of g_8C_1 ; therefore the only configurations based on this pattern are two invariant under G_{22} and two under G_{23} .

Proof of Theorem III. This theorem follows from the possibility of placing the odd counter in the unique central square (0, 0) of the network. The symmetry of configurations of $2r$ counters invariant under the 4- and 8-order groups and of $G_{21} = \{R^2\}$ is not affected, and yields a unique configuration of the same type with $2r + 1$ counters. Conversely all patterns involving $2r + 1$ counters in these groups must have the odd counter at (0, 0), the removal of which yields a unique configuration of the same type with $2r$ counters.

Manuscript received 9/9/58

Revised manuscript received 1/22/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A SYNTHESIS OF TWO FACTOR ANALYSES OF INTERMEDIATE ALGEBRA*

WILLIAM E. KLINE

BOARD OF EDUCATION
OF BALTIMORE COUNTY, MARYLAND

A battery of 18 tests of intermediate algebra and 20 reference tests was administered to two successive second-year algebra classes. Each battery was separately factor analyzed by Thurstone methods, and the two analyses were synthesized by the Tucker method. The five congruent factors obtained were identified as: Verbal Comprehension, Deductive Reasoning, Algebraic Manipulative Skill, Number Ability, and Adaptability to a New Task.

This study aims to explore intensively the small area of intermediate, or second-year, algebra by surveying the abilities employed in solving the variety of problems included in an intermediate algebra course. The data collected for the study provide an excellent opportunity to apply, and thus to test, Tucker's method [14] for synthesis of factor analysis studies.

The author's hypotheses before conducting the study included: (i) for the solution of algebra problems important basic abilities are inductive and deductive reasoning, rote memory, number, verbal comprehension, and spatial visualization; (ii) a basic ability of algebraic manipulative skill is involved in intermediate algebra; (iii) fluency of expression is required in the solution of statement problems; (iv) all abilities necessary for the solution of statement problems can be measured by multiple choice tests.

To test these hypotheses a battery of 38 tests was administered to each of two successive second-year algebra classes. Of these tests 18 were algebra tests and 20 were reference variables. Separate tables of intercorrelations were completed for each of the two years, and each was factor analyzed by the Thurstone [11] complete centroid method. The separate factor analyses were combined by Tucker's technique [14] for synthesis of factor analysis studies. The loadings of factors congruent to both analyses were determined. The axes were rotated in the congruent space to simple structure prior to factor interpretation.

To determine the efficacy of multiple choice answers for statement problems, two pairs of similar tests were constructed and included in the analysis. One of each pair used free answers and the other used multiple choice answers. The factor content of these pairs of similar tests was compared to determine whether or not different abilities were measured by the two methods of answering statement problems.

---

*This paper is a condensation of a thesis. The work was begun while the author was a Psychometrics Fellow of the Educational Testing Service. The work was further supported by Contract N6onr-270-20 of the Office of Naval Research and by Grant NSF G-642 of the National Science Foundation. The writer is indebted to Professors Harold Gulliksen and Ledyard R Tucker for their guidance throughout this study.

Description of the Variables

The following is a brief description of the 38 variables used in the study.* In the parentheses following the description of each reference variable, numbered 19–38, is the name of the factor which the author expected to be identified by the test.

1. *Fundamental Operations*: algebraic addition, subtraction, multiplication, and division of monomials and polynomials.
2. *Fractions*: reducing, adding, subtracting, multiplying, and dividing algebraic fractions; multiplying mixed expressions; simplifying complex fractions.
3. *Factoring*: all of the common types of factoring exercises.
4. *Quadratic Equations*: some solvable by factoring, others requiring the quadratic formula or method of completing the square.
5. *Radicals*: simplifying monomial radicals; fundamental operations of monomial and binomial radical expressions.
6. *Exponents*: problems based on the four basic laws of exponents; problems based on the meaning of the zero exponent, the negative exponent, and the fractional exponent.
7. *Binomial Theorem*: questions aimed to measure the student's thorough knowledge of the binomial theorem.
8. *Progressions*: questions testing the student's knowledge of arithmetic and geometric progressions.
9. *Use of Tables*: reading tables of common logarithms, trigonometric functions, and logarithms of trigonometric functions.
10. *Principles of Logarithms*: questions aimed at measuring the student's comprehension of logarithmic principles, the understanding of the definition of logarithm and its implications, and the four basic laws.
11. *Cartesian Graphs*: questions on elementary analytic geometry of the straight line and the conic sections.
12. *Simultaneous Equations*: pairs of equations, linear or quadratic.
13. *Informational Ability in Algebra*: aimed at measuring the student's familiarity with new verbal concepts introduced into mathematics at the intermediate algebra level.
14. *Theory of Quadratics*: questions on the nature of the roots, the discriminant, the relationship of the roots to the coefficients, and the parabolic graph of quadratic equations.
15. *Converting Statements to Symbols—A*: direct translation of words into algebraic symbols.
16. *Converting Statements to Symbols—B*: like Test 15.
17. *Word Problems—B*: writing the algebraic expressions for the given verbal phrases or the equations that would solve the stated problem.
18. *Word Problems—A*: like Test 17.
19. *Addition and Division*: addition of three one-digit or two two-digit numbers; division of two-digit or three-digit numbers by a number between 2 and 12, inclusive. Prepared by Educational Testing Service [3]. (Number)
20. *Subtraction and Multiplication*: subtraction of one-digit or two-digit numbers from

---

*Copies of the tests have been deposited as Document number 6006 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting $15.00 for photoprints, or $4.75 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

two-digit numbers; multiplication of two-digit or three-digit numbers by one-digit or two-digit numbers. Prepared by Educational Testing Service [3]. (Number)

21. *Opposites*: selecting from five words the one which is opposite in meaning to the given word. Similar to tests of Garrett [4], Harrell [6], Peterson [8], and Sisk [9]. (Verbal Comprehension)
23. *Similes*: patterned after Taylor's test [10]. The student finished incomplete similes with three different words or phrases. (Fluency of Expression)
24. *Inventive Opposites*: patterned after Thurstone's test [12]. The student responded with two words opposite in meaning to the given word. The initial letters of the correct responses were given. (Fluency of Expression)
25. *Letter-Blank Sentences*: sentences created by the subject, with the number of words equal to the given number of dashes and letters. Each word replacing a letter was required to begin with that letter. (Fluency of Expression)
26. *Figure Analogies*: similar to Army Air Force's Figure Analogies [1]. (Deductive Reasoning)
27. *Block Counting*: the student counted in a pile of blocks, pictured in the test, the number of shaded blocks that touched the numbered blocks. Prepared by Educational Testing Service [3]. (Space)
28. *Letter Series*: patterned after a Thurstone test [13]. The subject was required to indicate the fourth letter after the last letter in the given series. (Inductive Reasoning)
29. *Letter Sets*: similar to tests of the Thurstones [5, 13], Coombs [2], and Taylor [10]. The subject selected the one group of four letters that did not possess the characteristic common to the other four groups of letters in the item. (Inductive Reasoning)
30. *Family Trees*: similar to the Thurstones' Pedigrees [13]. The subject answered questions concerning relatives of certain members on the given family tree chart. (Inductive Reasoning)
31. *Following Directions*: similar to the Thurstones' test [13], in which the subject performed simple tasks on the paper with a pencil, following easy directions (Deductive Reasoning)
32. *Similar figures*: prepared by Educational Testing Service [3]. The subject selected the one of the four given figures which was a reversal. (Space)
33. *Tabular Completion*: similar to Thurstone's test [12]. The subject filled in missing data in tables of numerical data. (Deductive Reasoning)
34. *Word-Word*: similar to Garrett's tests [4]. A paired-associates test composed of common four-letter words. The subject was given four minutes to learn 20 pairs; immediately thereafter he was asked to complete each pair, when given one of the two words. (Rote Memory)
35. *Number-Number*: similar to Thurstone's test [12], like Test 34, but using pairs of two-digit numbers. (Rote Memory)
36. *Final Examination in Algebra*: a test of The Choate School. The same test was used for the two years of the study. (Grades)
37. *Second Semester Grade in Algebra*: for the student at Choate for the year of the study. (Grades)
38. *Final Grade in English*: for the student at Choate for the year of the study. (Grades)

Two Analyses and Their Synthesis

The battery of tests was administered to all students in second-year algebra in 1950–51 and 1951–52 at The Choate School, an independent boys' preparatory school in Wallingford, Connecticut. About 60 percent of these students had studied their first-year algebra at Choate; the others had begun their study of algebra in the school they attended prior to attending Choate.

The number of subjects included in the study was, by coincidence, 126 for each year. The tests were given in 19 testing periods, none longer than 40 minutes. Although the reference tests were given without advance notice, the algebra tests were announced at least five days in advance.

Except for differences in teaching methods used by the five teachers in each of the two school years, the 11 divisions of second-year algebra were equally prepared for the algebra tests, in that all students received similar daily assignments. In 1950–51 the textbook used was: Welchons, A. M. and Krickenberger, W. R., *Algebra Book Two*; Ginn, 1949. In 1951–52 the textbook was: Snader, Daniel W., *Algebra (Meaning and Mastery) Book Two*; Winston, 1950.

Instructions on the 18 algebra tests were brief, simple, and included the time limit, varying from 10 to 30 minutes. The least number of items in any test was 24. The author prepared all the items for these tests. The scores on the tests were converted into grades, which were a large part of the student's final record in algebra for the year's course. To achieve optimum performance from the students, the author informed them before the testing that final grades would depend upon their performance on these tests.

Instructions on the reference tests were often quite elaborate, usually including examples and practice problems. Time limits varied from 4 to 20 minutes. The brevity of the items enabled more to be administered in a shorter time than was true of the items on algebra tests. The number of items ranged from 26 to 100. Except for those five tests used with the permission of Educational Testing Service all the test items were original with the author. For motivation the students were informed at the beginning of the study that their scores on these tests would be used for guidance purposes.

Three of the reference variables were students' grades as reported at the end of the year—Final Examination in Algebra, Second Semester Grade in Algebra, and Final Grade in English. The data for the first year of the study have been labeled Study A; the second year's data carry the notation Study B.

The scores for the individuals' tests were intercorrelated separately for each year of the study. The two 38 × 38 matrices of intercorrelations are given in Table 1. Each of the matrices was factored by the complete centroid method, using the highest entry in a column as an estimate of the communality at each stage of the factoring. Distributions of the final residuals, after 12 factors were removed in each study, indicate that the common factor variance had been exhausted. The twelfth-factor residuals appear in Table A^*, and distributions of these final residuals are given in Table B. The pro-

---

*Copies of Tables A-M have been deposited as Document number 6007 with the
ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Wash-
ington 25, D. C. A copy may be secured by citing the Document number and by remitting
$2.50 for photoprints, or $1.75 for 35 mm. microfilm. Advance payment is required. Make
checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

Intercorrelation Matrices for 38 Variables

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1		31	28	35	32	29	25	25	24	26	22	37	19	09	21	22	24	18	21	23	21	23	12	16	-04	18	06	24	14	23	16	15	21	06	00	23	33	27
2	35		59	52	57	41	34	32	32	45	29	49	49	28	24	30	39	37	16	26	30	28	15	24	18	26	20	40	25	34	20	25	31	00	12	38	44	23
3	44	64		49	52	49	38	30	32	31	34	46	41	34	23	36	38	40	19	19	24	14	00	27	12	26	14	35	21	27	11	18	28	02	10	45	43	24
4	28	49	54		44	46	35	46	33	31	33	43	42	27	18	30	39	44	18	32	17	15	00	09	07	18	04	26	30	19	15	11	17	00	08	32	48	27
5	43	62	59	47		66	37	39	32	46	37	62	50	31	32	41	41	45	26	31	33	41	13	22	14	31	15	38	20	29	24	17	38	01	18	55	55	41
6	43	59	53	39	66		49	53	42	41	38	47	51	38	33	47	46	57	18	23	34	34	-02	28	13	51	12	39	13	33	31	37	44	25	16	63	68	56
7	30	28	26	29	36	36		36	36	39	17	32	31	27	21	26	36	29	20	18	19	24	-02	18	20	39	11	28	15	32	05	25	31	05	10	52	53	27
8	43	30	37	31	37	40	42		39	50	48	57	60	39	37	53	43	44	35	37	31	35	01	28	16	32	02	44	25	31	38	26	42	12	20	65	67	42
9	45	38	36	33	44	43	45	52		51	41	34	49	41	43	41	35	43	25	23	25	30	-01	12	08	37	22	44	27	31	34	33	43	05	-03	43	55	31
10	34	39	36	36	50	54	42	62	64		46	50	63	58	41	42	47	43	33	32	36	42	18	24	18	43	22	48	39	39	31	38	47	01	09	49	60	42
11	22	11	17	19	25	29	26	26	33	30		43	52	37	37	33	38	50	31	28	23	36	18	28	07	17	25	37	27	33	30	35	39	01	02	41	48	32
12	37	45	44	39	51	42	41	45	45	52	27		55	40	35	41	37	46	31	35	28	39	14	19	15	29	19	38	31	27	30	28	33	05	12	50	57	38
13	41	40	40	32	51	52	38	60	53	64	39	54		52	52	59	57	64	30	24	50	52	09	14	45	45	26	47	32	45	48	26	44	11	10	58	61	44
14	31	39	32	29	42	51	43	44	53	64	20	52	53		31	36	32	37	24	25	31	38	01	19	09	37	09	27	27	31	25	29	41	-01	08	44	47	38
15	45	29	38	22	40	41	33	53	46	56	31	49	68	47		68	52	52	28	25	50	48	21	29	13	34	25	40	23	44	58	20	34	18	07	46	41	43
16	39	44	38	29	48	51	33	49	50	65	20	53	69	54	78		60	64	16	14	47	47	10	33	16	44	19	39	18	40	44	25	35	23	15	52	45	42
17	39	35	31	24	45	46	31	51	47	55	36	40	55	42	55	54		77	21	14	51	46	12	38	14	39	14	40	21	38	36	19	34	21	15	51	50	41
18	41	39	31	26	44	47	37	56	51	56	32	44	72	53	67	68	68		22	21	54	50	05	46	14	48	26	45	23	36	42	25	42	21	05	51	57	50
19	33	08	17	20	18	15	11	43	35	33	13	36	31	23	27	20	32	35		82	31	28	20	18	21	04	-00	34	41	31	43	05	57	05	07	26	31	30
20	21	00	08	08	06	03	15	32	31	26	07	26	17	13	16	07	18	21	75		26	21	22	15	24	01	-04	38	41	23	36	02	54	07	10	19	36	26
21	23	21	22	18	22	32	15	43	34	50	32	29	64	34	67	61	50	59	25	12		69	13	55	20	36	24	30	22	47	56	08	36	08	07	29	32	59
22	29	14	18	11	27	28	21	54	36	55	27	32	64	39	61	58	51	58	32	17	72		23	49	30	34	28	34	30	66	52	19	43	12	14	36	37	53
23	11	07	08	05	16	07	09	15	16	23	07	11	28	17	19	20	33	25	08	11	24	34		11	37	-17	19	06	21	22	14	04	05	-12	04	07	02	07
24	21	30	28	14	33	31	26	46	40	50	12	38	59	37	61	68	41	57	29	11	55	59	18		17	34	23	34	19	38	33	11	35	18	11	30	28	39
25	26	26	30	21	36	29	31	34	40	44	16	41	36	32	45	41	40	38	45	37	37	39	29	48		07	13	17	35	33	22	08	16	08	07	18	04	12
26	32	27	28	20	51	42	31	31	45	45	34	35	42	39	45	49	42	54	15	00	26	29	09	43	30		31	48	05	36	34	46	33	43	30	47	52	41
27	19	05	07	08	24	16	23	28	28	28	27	26	27	19	23	19	23	19	09	-01	13	29	16	18	26	31		33	20	33	25	45	14	03	18	11	20	12
28	38	26	29	14	44	37	22	31	45	44	27	24	35	24	36	37	35	26	21	17	26	36	10	34	43	47	35		38	36	40	41	48	18	21	35	45	37
29	25	24	24	23	35	27	14	31	29	35	13	28	39	23	36	34	29	31	30	25	32	35	17	44	39	32	27	36		36	29	17	34	03	09	13	22	23
30	31	09	14	02	30	23	23	37	42	48	20	22	45	28	45	43	48	43	30	19	45	57	22	41	48	38	30	52	33		49	25	49	14	22	31	29	38
31	26	14	08	07	20	24	17	44	34	44	15	28	52	25	53	47	48	51	32	28	56	60	34	49	48	29	22	39	41	57		22	47	21	14	29	27	43
32	26	22	16	08	27	29	10	22	25	24	23	24	28	23	28	31	26	29	02	05	15	15	12	14	20	35	43	23	18	22	14		22	22	19	30	45	18
33	43	21	25	11	33	39	23	48	38	52	18	36	48	35	45	42	47	48	49	39	39	50	06	41	34	38	20	45	32	61	49	25		02	05	40	42	30
34	29	24	13	13	24	27	24	20	31	37	12	26	35	36	29	36	29	36	18	09	21	23	19	32	32	29	21	16	14	25	23	22	21		29	11	16	21
35	13	10	10	04	05	15	23	19	20	22	09	15	18	25	15	18	09	08	14	11	02	06	03	17	23	06	03	18	09	14	01	05	10	38		17	19	24
36	30	36	24	31	44	45	52	55	54	69	40	52	58	61	50	56	50	63	28	19	37	50	19	45	40	48	35	32	34	41	36	29	38	40	11		73	44
37	45	46	43	41	58	55	56	70	67	78	35	64	70	68	62	68	58	69	39	27	48	53	23	50	45	45	27	32	38	40	40	26	43	42	27	78		60
38	23	39	29	27	32	36	32	53	50	61	15	46	54	50	52	52	42	51	37	25	54	53	20	50	39	24	18	22	43	36	44	12	23	23	14	55	65	

Decimal points have been omitted.
Study A correlations are below the diagonal.
Study B correlations are above the diagonal.

TABLE 2

Projections of the Test Vectors upon the Orthogonal
Centroid Reference Vectors--Study A

F_{jnA}

Test	FACTORS												h^{2*}
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
1	.55	.14	-.13	-.17	.13	-.14	-.10	.17	-.12	.08	-.04	-.10	.48
2	.52	.38	.28	-.30	.17	.12	-.21	-.10	.12	.04	-.03	.03	.70
3	.51	.32	.20	-.40	.23	-.05	-.16	.02	-.08	.07	-.12	.06	.67
4	.42	.32	.23	-.31	.01	-.11	-.04	-.09	.06	.05	-.11	.03	.47
5	.65	.37	.06	-.20	.26	.06	.05	-.10	.03	.12	.03	.11	.72
6	.64	.35	.12	-.11	.20	.09	-.08	.11	.12	-.12	.07	.08	.67
7	.51	.27	.05	.08	-.22	.03	.09	.05	-.08	.10	-.12	-.08	.44
8	.71	-.06	.08	-.04	-.11	-.18	.16	.13	-.11	-.10	-.09	-.18	.67
9	.71	.16	-.10	.03	-.14	.02	.11	.16	.13	.12	-.12	-.11	.66
10	.81	.05	.08	.06	-.15	.13	.12	.09	.11	-.12	.01	.11	.77
11	.40	.14	-.08	.19	.09	-.34	.14	.12	.09	.06	-.12	.09	.41
12	.66	.18	.13	-.13	-.13	-.09	-.04	-.09	-.14	-.02	.13	.04	.58
13	.81	-.09	.19	.12	.10	-.12	-.05	.04	-.02	-.11	.05	.05	.76
14	.66	.20	.18	.13	-.24	.10	-.05	.06	.08	-.12	.12	.05	.64
15	.76	-.21	.17	.13	.16	-.06	-.11	.10	-.14	.07	.05	-.13	.76
16	.77	-.07	.28	.18	.14	.13	-.14	.07	-.08	.08	.17	-.08	.82
17	.71	-.11	-.03	.08	..09	-.07	-.08	.08	.20	.08	-.11	.06	.61
18	.77	-.15	.15	.19	.07	-.12	-.08	.08	.16	.19	.17	-.05	.80
19	.48	-.32	-.28	-.47	-.34	-.17	.07	.07	-.04	.12	.19	.10	.85
20	.33	-.29	-.32	-.43	-.44	-.14	.03	.11	.09	.07	.16	.04	.75
21	.62	-.44	.25	.15	.18	-.06	-.04	.05	.05	-.05	-.12	.11	.73
22	.67	-.45	.13	.18	.09	-.05	.14	-.02	-.12	-.08	-.11	.16	.79
23	.29	-.17	-.06	.15	-.04	-.06	-.20	-.22	.11	.03	-.20	.12	.30
24	.65	-.29	.17	.06	.05	.27	.04	-.13	-.19	.15	.13	-.08	.71
25	.61	-.14	-.23	-.10	-.12	.14	-.12	-.19	-.10	.14	-.10	.12	.59
26	.58	.18	-.12	.19	.25	.11	.17	-.02	.07	.22	.21	-.06	.62
27	.37	.11	-.30	.23	.09	-.15	.16	-.31	-.16	-.17	-.05	-.09	.51
28	.54	.05	-.36	-.04	.26	.24	.19	.07	-.06	.03	-.17	.05	.63
29	.50	-.13	-.09	-.16	.10	.13	.11	-.26	.04	-.06	.07	-.11	.43
30	.59	-.27	-.31	.14	.13	.22	.14	.08	-.07	-.06	-.15	.11	.67
31	.59	-.49	-.11	.04	.10	.11	-.03	-.08	.07	-.06	-.09	-.04	.65
32	.37	.18	-.24	.18	.21	-.18	-.09	-.14	.04	-.17	.18	-.13	.44
33	.62	-.19	-.22	-.10	.13	.03	.12	.27	-.07	-.13	.19	.20	.68
34	.44	.13	-.14	.26	-.17	.08	-.30	-.06	-.10	.04	.07	.07	.45
35	.24	.12	-.12	.06	-.28	.22	-.18	.17	-.25	.02	-.05	-.04	.35
36	.74	.13	.09	.23	-.19	-.05	.24	-.15	.08	.04	.09	.08	.79
37	.86	.14	.19	.07	-.25	-.06	.07	.05	.06	-.02	.04	-.04	.88
36	.66	-.15	.28	-.04	-.23	.08	.05	-.14	.14	-.13	-.11	-.14	.69

*h^2 = the communality

TABLE 3

Projections of the Test Vectors upon the Orthogonal
Centroid Reference Vectors--Study B

F_{jMB}

Test	FACTORS												h^{2*}
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
1	.38	.10	-.14	-.06	.09	-.15	.09	.06	.06	-.07	-.19	-.24	.32
2	.58	.22	-.23	-.27	.24	-.04	.14	.04	.02	.07	.07	.08	.61
3	.54	.32	-.18	-.20	.18	-.15	.15	.04	-.18	.11	.06	.15	.62
4	.50	.35	-.25	-.15	.07	-.21	.05	.22	.10	.13	.11	-.16	.62
5	.65	.23	-.13	-.13	.13	-.19	.04	-.11	.07	.05	-.17	.11	.62
6	.70	.31	.10	.15	.13	-.19	.14	-.15	-.06	.06	-.21	.02	.77
7	.50	.26	-.05	.12	.16	-.07	.00	-.24	-.20	-.09	.11	-.14	.50
8	.68	.18	-.04	.16	-.18	-.07	-.21	.05	.14	.05	-.05	.09	.64
9	.58	.18	-.05	.13	-.11	.29	.10	.09	-.10	-.02	.04	-.18	.55
10	.70	.13	-.16	.06	-.09	.29	-.05	-.11	.11	.04	.13	-.03	.67
11	.58	.09	-.10	-.07	-.15	.17	.10	.10	.12	-.18	-.14	.05	.50
12	.66	.22	-.21	-.10	.04	-.03	-.06	.07	.25	.07	-.15	.10	.65
13	.77	.15	.07	-.12	-.20	.15	.02	.05	.10	.17	.05	.12	.76
14	.55	.16	-.04	.13	-.21	.23	.05	-.14	.09	.19	.12	.07	.53
15	.63	-.14	.23	-.10	-.23	.15	-.20	.14	-.06	.03	-.13	-.10	.65
16	.67	.06	.29	-.15	-.15	.05	-.19	.14	-.12	.17	-.12	.05	.70
17	.68	.11	.22	-.22	-.13	-.11	-.17	.09	-.19	-.03	.11	-.07	.69
18	.73	.13	.27	-.22	-.20	-.08	.04	.19	-.15	-.14	.10	.06	.79
19	.49	-.39	-.47	.26	-.23	-.18	-.04	.11	-.05	-.09	-.04	.05	.79
20	.48	-.33	-.58	.27	-.16	-.26	-.04	.15	.03	-.04	-.04	.05	.87
21	.61	-.31	.26	-.22	-.23	-.15	.22	-.10	.08	.07	.15	.03	.75
22	.66	-.33	.20	-.21	-.09	.03	.10	-.26	.12	.04	-.09	-.09	.75
23	.18	-.29	-.19	-.31	.04	.08	-.30	-.12	.17	-.21	-.08	.08	.45
24	.49	-.24	.21	-.15	-.03	-.14	.16	-.08	-.09	-.09	.14	.14	.47
25	.29	-.32	-.16	-.11	.20	.06	-.29	-.25	-.14	.10	.07	.08	.46
26	.57	.10	.43	.31	.26	.07	.14	.03	-.12	.09	.13	.04	.74
27	.33	-.14	.17	-.14	.31	.38	.09	.07	.13	-.26	.10	.16	.55
28	.64	-.07	-.06	.11	.14	.13	.10	.27	-.08	-.02	.09	.10	.57
29	.44	-.26	-.35	-.06	.05	.14	-.05	.07	.03	.08	.18	-.06	.46
30	.62	-.32	.05	-.12	.10	.17	.09	-.13	-.06	.11	-.07	-.12	.60
31	.58	-.40	.15	.03	-.15	.07	.08	.20	.03	.16	-.18	.05	.66
32	.43	.11	.12	.23	.34	.37	.03	.07	.12	-.21	-.04	.05	.59
33	.62	-.16	-.19	.22	-.18	.09	.21	-.06	-.24	-.08	-.12	.14	.68
34	.21	-.10	.37	.23	.23	-.13	-.14	.24	-.11	.13	-.07	-.05	.43
35	.23	-.11	.16	.17	.32	-.12	-.15	.04	.15	.15	.05	.11	.32
36	.70	.31	.08	.14	-.08	-.06	-.20	-.20	-.09	-.04	-.09	.16	.75
37	.77	.35	.03	.27	-.05	-.12	-.04	-.05	.18	-.16	.07	-.03	.87
38	.64	-.10	.21	.10	-.08	-.22	.10	-.14	.18	.12	.06	-.10	.62

*h^2 = the communality

jections of the 38 test vectors upon the 12 orthogonal centroid reference vectors for Study A are listed in Table 2. Those for Study B are listed in Table 3.

Following Tucker's notation [14] the reference-factor matrices in Table 2 and 3 are labeled F_{imA} and F_{jMB}, respectively. The subscripts, m and M, designate the reference factors. In synthesizing the two analyses two transformation matrices, T_{mrA} and T_{MrB}, must be found so that only negligible differences in the matrices F_{jrA} and F_{jrB} will result when

$$F_{jrA} = F_{jmA}T_{mrA} \quad \text{and} \quad F_{jrB} = F_{jMB}T_{MrB} .$$

Only one of the 38 variables was not considered as an overlap test. Test 1, *Fundamental Operations,* was given without earlier preparation in Study A; whereas in Study B the students were given assigned work in this topic prior to the test. For this reason it was eliminated from the test of congruence.

Table 4 lists the means and the standard deviations of the overlap

TABLE 4

Means and Standard Deviations of the Overlap Variables

Test	M_A	M_B	σ_{jA}	σ_{jB}
2	20.42	21.00	4.74	4.59
3	17.41	18.88	3.81	3.60
4	28.11	27.48	5.78	7.49
5	17.48	20.52	5.10	3.80
6	20.33	20.90	4.68	5.78
7	17.13	20.44	5.22	3.86
8	9.60	11.15	3.79	3.90
9	11.11	8.01	5.21	4.33
10	14.55	13.60	5.25	4.98
11	10.85	13.01	3.46	4.55
12	16.63	17.56	4.25	4.78
13	16.78	17.00	6.94	6.53
14	11.02	10.67	3.68	3.26
15	17.77	19.60	3.96	2.94
16	16.54	18.00	5.07	3.47
17	12.90	14.60	4.12	3.93
18	9.99	11.29	4.53	4.75
19	37.37	38.56	7.17	8.57
20	30.49	31.31	7.24	8.36
21	22.22	25.35	7.99	8.14
22	19.75	22.49	5.62	5.42
23	24.42	24.88	7.53	7.09
24	37.77	41.69	5.55	4.87
25	16.77	16.36	4.73	4.61
26	25.82	26.92	4.90	4.93
27	15.69	17.43	6.05	6.71
28	9.06	9.90	3.55	4.10
29	14.13	15.42	3.24	3.39
30	16.30	18.60	5.48	5.46
31	17.47	18.81	3.77	4.00
32	17.42	18.84	7.28	8.26
33	22.31	25.17	10.53	9.79
34	16.33	17.87	7.16	8.15
35	9.78	9.45	4.86	4.05
36	75.39	76.91	14.83	13.89
37	73.92	76.13	9.39	8.33
38	75.44	77.09	6.48	7.01

M_A = Mean score in Study A

M_B = Mean score in Study B

σ_{jA} = Standard Deviation of the scores in Study A

σ_{jB} = Standard Deviation of the scores in Study B

TABLE 5

Latent Roots

Study A		Study B	
Axis	A	Axis	B
p-1	14.4720	P-1	12.2873
p-2	2.3614	P-2	2.3003
p-3	1.1513	P-3	2.0106
p-4	1.5515	P-4	1.2271
p-5	1.2286	P-5	0.9263
p-6	0.7608	P-6	1.2936
p-7	0.5706	P-7	0.6672
p-8	0.6358	P-8	0.7549
p-9	0.4064	P-9	0.5348
p-10	0.3663	P-10	0.4849
p-11	0.5427	P-11	0.3932
p-12	0.2553	P-12	0.2663

TABLE 6

Coefficients of Congruence (arranged in order of size)

ϕ_r^2	ϕ_r
.9860	.9930
.9379	.9685
.9151	.9566
.8570	.9257
.7300	.8544
.3777	.6146
.2754	.5248
.2120	.4604
.0084	.0917

variables for each study. In order that a common unit of measurement would be used for each test in both studies, adjustments were made on the standard deviations, computed by the formulas

$$d_{jA} = \frac{\sigma_{jA}}{\frac{1}{2}(\sigma_{jA} + \sigma_{jB})} \quad \text{and} \quad d_{jB} = \frac{\sigma_{jB}}{\frac{1}{2}(\sigma_{jA} + \sigma_{jB})} .$$

Reference-factor matrices for the tests with adjusted units of measurement were computed by

$$F_{JmA} = D_{JA}F_{jmA} \quad \text{and} \quad F_{JMB} = D_{JB}F_{jMB} .$$

The subscript J designates the tests after the adjustment of the standard deviations. Matrices F_{JmA} and F_{JMB} are given in Tables C and D.

The latent roots and latent vectors of the matrices $F'_{JmA}F_{JmA}$ and $F'_{JMB}F_{JMB}$ were computed on the IBM 701. The latent roots for both matrices are shown in Table 5. The latent vectors for Study A are given in matrix Λ_{mpA} in Table E, and the latent vectors for Study B are given in matrix Λ_{MPB} in Table F.

Three principal axes in each study (p–9, p–10, p–12, P–10, P–11, and P–12), corresponding to the lowest latent roots in Table 5, were discarded because it is necessary to exclude from the congruent space those dimensions into which the overlap tests have small projections. As yet no rule has been formulated to determine which principal axes should be retained and which should be relegated to the noncongruent space. In this study the author arbitrarily chose 0.5 as the critical value of the latent root.

The Gramian matrix H_A is given in Table G. The latent roots of this matrix are the squares of Tucker's coefficient of congruence, ϕ_r .

$$H_A = GG',$$

in which

$$G = T'_{mpA}(F'_{JmA}F_{JMB})T_{MPB} ,$$

where

$$T_{mpA} = \Lambda_{mpA}\beta_A^{-1/2} \quad \text{and} \quad T_{MPB} = \Lambda_{MPB}\beta_B^{-1/2}.$$

β_A is a diagonal matrix containing the latent roots of $F'_{JmA}F_{JmA}$, and β_B is a diagonal matrix containing the latent roots of $F'_{JMB}F_{JMB}$. The latent vectors of H_A are given in matrix Λ_{rA} in Table H.

Table 6 gives the values of ϕ_r for the nine factors remaining in each study. As yet no criteria have been developed as to what minimal value of ϕ_r may be acceptable for indicating congruence. Because it is similar to the formula for the product moment correlation between the loadings on the factor r for Studies A and B, those values in the range acceptable for reliability of good tests might be considered as being indicative of considerable

TABLE 7

Transformation Matrix -- Study A

T_{mrA}

Centroid factors	Rotated factors A	B	C	D	E
i	.6818	.0217	.0318	-.0919	.0598
ii	-.2379	.5222	-.0426	.0510	.5645
iii	-.2146	.5058	.3699	.3226	-.5724
iv	.3210	-.1296	-.3033	.9286	.1934
v	.1760	-.0243	.1969	.3012	.1063
vi	-.2682	-.3125	.1907	.0006	.2112
vii	.0903	.4910	-.4520	-.0450	-.2472
viii	.0721	.4432	-.4722	.0118	-.4261
ix	.0946	.2347	.0438	.0028	-.1319
x	-.0646	-.1297	-.0001	-.0974	.1331
xi	-.0437	.0684	-.5727	-.1007	.0565
xii	.0739	-.0546	.1213	-.1105	-.0233

TABLE 8

Transformation Matrix -- Study B

T_{MrB}

Centroid factors	Rotated factors A	B	C	D	E
I	.7386	.0216	.0397	-.0893	.0476
II	-.2108	.7775	.0053	.1937	.1150
III	.3236	-.1344	-.0575	.7316	-.0499
IV	-.0707	.1194	-.7326	-.2608	.1122
V	-.2733	- -.2472	.3249	-.1919	.8580
VI	.3767	-.0499	-.4148	.3881	.2324
VII	-.3836	.3112	.0219	-.0640	-.2498
VIII	-.1546	-.1314	-.0555	-.0909	.1282
IX	-.2797	.1437	.1981	.2067	-.2112
X	-.1449	-.0151	.2275	-.0447	-.0868
XI	-.0541	-.0011	.0873	-.0604	.0983
XII	.0555	-.0579	.0064	-.0313	.1190

TABLE 11

γ_{rs}

	a	b	c	d	e
A	.490	.512	.121	.269	.264
B	-.406	.581	.427	-.131	-.585
C	.496	-.534	.779	-.346	.067
D	.275	.139	-.348	-.851	.006
E	-.522	.309	.274	-.258	.764

TABLE 9 Study A

Factor Loadings of Overlap Tests on the
Five Factors in the Congruent Space
$F_{JrA} = F_{JmA}T_{mrA}$

	A	B	C	D	E
2	.10	.23	.42	-.16	.16
3	.14	.24	.38	-.27	.08
4	.08	.27	.25	-.20	.03
5	.35	.24	.20	-.18	.29
6	.29	.26	.09	-.04	.11
7	.29	.22	-.05	.00	.14
8	.50	.21	-.10	-.08	-.19
9	.51	.19	-.13	-.10	.09
10	.53	.17	-.06	-.04	-.05
11	.40	.20	-.14	.12	.02
12	.30	.13	.04	-.17	.08
13	.64	.10	.04	.13	-.12
14	.37	.22	-.06	.06	.04
15	.67	-.06	.04	.18	-.14
16	.59	.00	.06	.25	-.04
17	.61	-.03	.07	.02	-.03
18	.61	.05	-.08	.14	-.11
19	.30	-.13	-.18	-.67	-.15
20	.18	-.11	-.22	-.63	-.15
21	.58	-.09	.19	.20	-.35
22	:65	-.14	.06	.15	-.29
23	.31	-.32	.27	.08	.09
24	.41	-.25	.10	.03	-.05
25	.38	-.40	.18	-.28	.22
26	.45	.06	-.21	.13	.29
27	.38	-.14	-.09	.14	.34
28	.39	-.12	-.02	-.11	.24
29	.30	-.15	.10	-.19	.07
30	.57	-.30	-.07	.01	.07
31	.53	-.37	.12	-.02	-.11
32	.36	-.04	-.13	.14	.32
33	.55	-.02	-.27	-.23	-.09
34	.27	-.23	-.01	.09	.32
35	.04	-.15	-.07	-.07	.20
36	.54	.19	-.15	.10	.08
37	.54	.29	-.07	-.02	-.05
38	.34	.07	.20	-.05	-.20

TABLE 10 Study B

Factor Loadings of Overlap Tests on the
Five Factors in the Congruent Space
$F_{JrB} = F_{JMB}T_{MrB}$

	A	B	C	D	E
2	.17	.16	.36	-.19	.21
3	.15	.22	.29	-.21	.18
4	.01	.34	.34	-.25	.06
5	.25	.17	.23	-.12	.06
6	.40	.32	.03	-.08	.09
7	.27	.18	-.02	-.12	.18
8	.50	.17	-.09	-.06	-.06
9	.43	.18	-.22	.01	.03
10	.54	.17	-.10	-.02	.03
11	.50	.17	-.08	.04	-.09
12	.33	.20	.20	-.12	.05
13	.60	.15	.05	.14	-.10
14	.42	.22	-.14	.05	-.11
15	.66	-.15	-.07	.18	-.11
16	.58	-.05	.03	.18	-.03
17	.64	.00	.13	.11	-.03
18	.67	.07	.09	.18	-.10
19	.32	-.18	-.20	-.62	-.18
20	.17	-.12	-.11	-.72	-.13
21	.52	-.14	.21	.12	-.35
22	.62	-.18	.16	.16	-.22
23	.29	-.29	.19	-.03	.05
24	.41	-.15	.13	.03	-.08
25	.38	-.37	.15	-.19	.23
26	.43	.02	-.16	.13	.30
27	.38	-.22	.03	.22	.36
28	.46	-.09	-.09	-.19	.26
29	.29	-.20	.07	-.30	.10
30	.57	-.25	.08	.00	.04
31	.54	-.29	-.07	.01	-.16
32	.37	.02	-.24	.13	.45
33	.51	.01	-.28	-.25	-.08
34	.21	-.25	-.07	.04	.25
35	.09	-.18	.09	-.04	.25
36	.60	.22	-.10	.02	.05
37	.42	.34	-.12	-.03	.02
38	.42	.03	.09	.02	-.20

congruence. The five factors having coefficients of congruence above .80 were retained in the congruent factor space; the remaining four were relegated to the noncongruent space.

Tables 7 and 8 show the matrices T_{mrA} and T_{MrB} which transform F_{JmA} and F_{JMB} into two matrices in which the differences are negligible for the overlapping tests.

$$T_{mrA} = \gamma_{mrA} D_r \quad \text{and} \quad T_{MrB} = \gamma_{MrB} D_r \ ,$$

in which

$$\gamma_{mrA} = T_{mpA} \Lambda_{rA} \quad \text{and} \quad \gamma_{MrB} = (T_{MPB} G')(\Lambda_{rA}) \phi_r^{-1} ,$$

where D_r is a diagonal matrix of the reciprocal of the square root of the average of the sums of the squares for entries in corresponding columns of γ_{mrA} and γ_{MrB} .

The matrix, $F_{JrA} = F_{JmA} T_{mrA}$, shown in Table 9, represents the factor loadings of the overlap tests in Study A on the five factors which are in the five-dimensional space congruent with Study B. The factor loadings of the overlap tests in Study B on the same five factors are shown in matrix $F_{JrB} = F_{JMB} T_{MrB}$ in Table 10.

Factor loadings for each pair of factors in the matrices F_{JrA} and F_{JrB} were plotted on orthogonal axes. After ten rotations the composite result of the transformation matrices, γ_{rs} , is given in Table 11. Tables 12 and 13 show, respectively, T_{msA} and T_{MsB} , the transformation matrices for converting the factor loadings on the rotated oblique factor axes.

TABLE 12

Transformation Matrices for Converting Factor Loadings on Orthogonal Axes to Factor Loadings on Rotated Oblique Axes

T_{msA}

	A	B	C	D	E
i	.2920	.3406	.1932	.2507	.1835
ii	-.6468	.3751	.3490	-.3310	.0516
iii	.2673	-.1415	.2450	-.4089	-.6524
iv	.2195	.4276	-.6126	-.6820	.2510
v	.2269	.0443	.1038	-.3252	.1341
vi	-.0205	-.3455	.0472	-.1643	.2446
vii	-.2694	.4765	-.2151	.2357	-.4128
viii	-.1572	.4048	-.3406	.2424	-.5106
ix	.0435	.1176	.1274	.0121	-.1796
x	-.0773	-.0785	.0083	.0520	.1367
xi	-.4006	.3175	-.4355	.2683	-.0405
xii	.1029	-.0789	.1315	.0919	.0351

TABLE 13

Transformation Matrices for Converting Factor Loadings on Orthogonal Axes to Factor Loadings on Rotated Oblique Axes

T_{MsB}

	A	B	C	D	E
I	.3318	.3614	.2034	.2653	.1888
II	-.4341	.3921	.3218	-.3831	-.3600
III	.4225	.1989	-.3882	-.5237	.1081
IV	-.5918	.4110	-.4766	.4445	-.0457
V	-.3828	-.2125	.4877	-.2284	.6400
VI	-.0159	.4968	-.4337	-.1498	.2400
VII	-.1957	-.1102	.0672	-.0352	-.4045
VIII	-.1456	-.0961	-.0601	.0422	.1109
IX	.0718	-.1963	.0609	-.3065	-.2605
X	.0831	-.2309	.1699	-.0597	-.0691
XI	-.0520	-.0515	.1276	-.0201	.0572
XII	-.0172	.0231	.0357	.0175	.1194

TABLE 14

Factor Loadings on Rotated Congruent Factors--Study A

$F_{JsA} = F_{JmA}T_{msA}$

	A	B	C	D	E
1	.03	.24	.19	.25	.11
2	.04	-.02	.63	-.06	.03
3	.04	.00	.62	.08	-.02
4	-.01	.05	.47	.07	-.08
5	-.03	.27	.51	.08	.17
6	.01	.27	.30	.02	.01
7	-.05	.34	.15	.03	.04
8	.19	.35	.06	.28	-.13
9	.03	.45	.12	.24	.07
10	.18	.37	.12	.21	.00
11	.06	.41	-.01	.02	.00
12	.03	.20	.24	.19	.05
13	.41	.33	.09	.07	.02
14	.06	.36	.10	.03	-.01
15	.50	.26	-.02	.06	.09
16	.42	.28	.02	-.08	.10
17	.38	.24	.12	.14	.13
18	.34	.40	-.06	.03	.04
19	-.01	.02	.04	.82	.01
20	-.08	.01	.00	.80	-.02
21	.67	.06	.01	.02	-.04
22	.62	.14	-.07	.13	.03
23	.40	-.13	.13	-.07	.30
24	.39	.00	-.01	.11	.19
25	.26	-.11	.21	.30	.44
26	-.03	.47	-.06	.00	.25
27	.06	.33	-.05	-.05	.37
28	.07	.20	.10	.18	.30
29	.17	.01	.16	.23	.20
30	.34	.17	-.11	.21	.32
31	.54	-.05	-.02	.21	.23
32	.00	.35	-.05	-.06	.30
33	.13	.34	-.16	.44	.06
34	.08	.12	-.02	-.06	.39
35	-.08	.02	-.04	.07	.21
36	.10	.50	.02	.08	.07
37	.13	.46	.16	.18	-.06
38	.34	.04	.23	.11	-.08

TABLE 15

Factor Loadings on Rotated Congruent Factors--Study B

$F_{JsB} = F_{JMB}T_{MsB}$

	A	B	C	D	E
1	.01	.02	.26	.09	-.03
2	.03	.02	.57	.01	.11
3	-.03	.07	.54	.05	.05
4	-.06	.00	.60	.05	-.12
5	.10	.10	.40	.06	.02
6	.01	.38	.30	.11	-.01
7	-.07	.28	.21	.12	.09
8	.16	.36	.07	.23	-.02
9	.01	.44	-.05	.17	.01
10	.13	.43	.08	-.06	.05
11	.20	.36	.04	-.02	-.04
12	.12	.17	.39	.09	.01
13	.36	.34	.11	.04	-.01
14	.11	.38	-.02	.13	-.09
15	.46	.27	-.15	.10	.15
16	.40	.26	.00	.01	.14
17	.43	.26	.15	.05	.14
18	.46	.32	.10	.02	.06
19	.05	.02	-.03	.81	.03
20	-.06	-.07	.11	.80	.00
21	.64	-.02	.04	.08	-.03
22	.63	.08	.01	.06	.10
23	.33	-.11	.10	.06	.25
24	.39	.02	.07	.09	.13
25	.24	-.05	.16	.22	.43
26	.00	.42	-.04	-.02	.27
27	.16	.19	.00	-.17	.44
28	.03	.27	.09	.28	.31
29	.13	-.02	.17	.34	.23
30	.40	.11	.05	.16	.29
31	.45	.10	-.19	.27	.16
32	-.16	.47	-.06	-.05	.35
33	.08	.35	-.10	.50	.04
34	.05	.08	-.10	.02	.33
35	.02	-.04	.10	-.01	.27
36	.14	.50	.10	.14	.05
37	-.01	.46	.13	.14	-.07
38	.36	.12	.08	.12	-.05

TABLE 16

Correlations of the Congruent Factors
Study A

	A	B	C	D	E
A	1.0000				
B	.3742	1.0000			
C	.1074	.4751	1.0000		
D	.2352	.0851	.0419	1.0000	
E	.0120	.2255	.0930	.1570	1.0000

TABLE 17

Correlations of the Congruent Factors
Study B

	A	B	C	D	E
A	1.0000				
B	.2323	1.0000			
C	.2556	.5012	1.0000		
D	.2104	.0489	.1992	1.0000	
E	-.1133	-.0758	-.1118	-.0708	1.0000

$$T_{msA} = \gamma_{msA} D_s \quad \text{and} \quad T_{MsB} = \gamma_{MsB} D_s \ ,$$

where $$\gamma_{msA} = T_{mrA} \gamma_{rs} \ , \quad \text{and} \quad \gamma_{MsB} = T_{MrB} \gamma_{rs} \ ,$$

and where D_s is a diagonal matrix of reciprocal square roots of mean sums of squares for entries in corresponding columns of γ_{msA} and γ_{MsB} .

Matrices F_{JsA} and F_{JsB} , shown in Tables 14 and 15, contain the loadings of the tests on the rotated factors obtained by

$$F_{JsA} = F_{JmA} T_{msA} \quad \text{and} \quad F_{JsB} = F_{JMB} T_{MsB} \ .$$

Tables 16 and 17 show the correlations between the congruent factors. Tables J and K include the direction cosines for the noncongruent factors. Tables L and M list factor loadings on these seven factors of Study A and B, respectively. These were determined by methods detailed in Tucker's paper [14].

Interpretation of the Results

Identification of the Factors

The five factors found in the congruent factor space have been labeled: Factor A, Verbal Comprehension; Factor B, Deductive Reasoning; Factor C, Algebraic Manipulative Skill; Factor D, Number Ability; Factor E, Adaptability to a New Task. All tests with factor loadings of .30 or above in either study are listed for each of the factors.

Factor A—Verbal Comprehension

Test no.	Algebra Tests	Study A Factor Loading	Study B Factor Loading
13	Informational Ability	.41	.36
15	Statements to Symbols—A	.50	.46
16	Statements to Symbols—B	.42	.40
17	Word Problems—B	.38	.43
18	Word Problems—A	.34	.46
	Reference Tests		
21	Opposites	.67	.64
22	Vocabulary Completion	.62	.63
23	Similes	.40	.33
24	Inventive Opposites	.39	.39
30	Family Trees	.34	.40
31	Directions	.54	.45
38	Final Grade in English	.34	.36

Factor A (Verbal Comprehension) had its highest loading in the two reference tests of vocabulary. Loadings above .30 appeared in all the reference tests that were largely composed of words except the paired-associates test.

In the algebra tests, Factor *A* was evident only in the four statement problem tests and in the test of informational ability.

Tests 15 and 16 involved the converting of verbal statements into algebraic symbols. Since the words were not unusual and were, in the author's opinion, thoroughly familiar to all the students, the large loading of these tests on Factor *A* suggest that this factor is somewhat contaminated with a reasoning element. Perhaps the factor could have been named "Verbal Reasoning."

Factor B—Deductive Reasoning

Test no.	Algebra Tests	Study A Factor Loading	Study B Factor Loading
6	Exponents	.27	.38
7	Binomial Theorem	.34	.28
8	Progressions	.35	.36
9	Use of Tables	.45	.44
10	Principles of Logarithms	.37	.43
11	Cartesian Graphs	.41	.36
13	Informational Ability	.33	.34
14	Theory of Quadratics	.36	.38
18	Word Problems—A	.40	.32
	Reference Tests		
26	Figure Analogies	.47	.42
27	Block Counting	.33	.19
32	Similar Figures	.35	.47
33	Tabular Completion	.34	.35
36	Final Examination in Algebra	.50	.50
37	Second Semester Algebra Grade	.46	.46

The nine algebra and six reference tests in which Factor *B* had loadings above .30 indicate that it is a factor of Deductive Reasoning. Tests 26 and 33 are similar to tests which have been used earlier in factor studies to identify a factor of deductive reasoning, defined here as the ability to reason from the general to the specific—to apply a principle or a rule to a specific problem when the principle or rule has been previously given, explained, and understood by the subject.

Factor C—Algebraic Manipulative Skill

Test no.	Algebra Tests	Study A Factor Loadings	Study B Factor Loadings
2	Fractions	.63	.57
3	Factoring	.62	.54
4	Quadratic Equations	.47	.60
5	Radicals	.51	.40
6	Exponents	.30	.30
12	Simultaneous Equations	.24	.39

Factor C (Algebraic Manipulative Skill) appeared only in those algebra tests composed of material that had been studied in elementary algebra and reviewed in the intermediate algebra course before the tests were given. Its highest loadings occurred in those tests which required a maximum of algebraic manipulation. The Final Examination in Algebra was in part composed of problems that required manipulation, but these problems constituted only 32 percent of the total score. The other variables on which a larger loading may have been expected were Test 1, Fundamental Operations, which was a very easy test, and variable 37, Second Semester Grade in Algebra. The second semester in each year of the study was devoted to the topics of logarithms, numerical trigonometry, progressions, and the binomial theorem. These topics involve very little algebraic manipulation in comparison with the topics taught during the first semester.

That all the tests with significant loadings on Factor C represent topics taught in elementary algebra suggests the possibility that this factor is one of previous training. Evidence to the contrary exists in the factor loadings of Tests 15–18, which represent topics usually included in elementary algebra. Since these tests had very small loadings on Factor C, it is believed the element of previous training is not important in this factor.

For all the tests having loadings above .30 on Factor C the directions were very concise, e.g., "simplify," "evaluate," "solve," "perform the indicated operations," etc. There is good evidence that the student was not required to change set in the course of the test; it is possible that this factor represents "an ability to maintain a fixed set." The author's thorough study of the other tests leads him to believe that frequent changes of set were demanded in each. The possibility exists, too, that the factor may involve both the ability to maintain a fixed set and algebraic manipulative skill. Further testing and analysis is necessary to determine the identity of this factor. Its high correlation of .48 with Factor B, as shown in Table 16, would suggest that some element of deductive reasoning may be involved in Factor C.

Factor D—Number Ability

Test no.	Reference Tests	Study A Factor Loading	Study B Factor Loading
19	Addition and Division	.82	.81
20	Subtraction and Multiplication	.80	.80
25	Letter-Blank Sentences	.30	.22
29	Letter Sets	.23	.34
33	Tabular Completion	.44	.50

The very high loadings of Factor D on the reference tests 19 and 20, incorporated in the analysis to aid in locating the number factor, provided the name for this factor. Test 33 also required a considerable amount of simple arithmetic computation. Test 25, on the other hand, contained no numbers, requiring only the writing of a sentence of a given *number* of words. It is difficult to understand why the old concept of the factor of number ability played an important role in scoring high on this test; however, it might have been that a sense of number aided in composing sentences of the required number of words more rapidly. Test 29 contained no numbers and required no arithmetic computation. The subject was required to determine the characteristic that is common to four of the given five groups of four letters each. For this reason Factor *D* may best be described as a factor of number skill contaminated by some ability involving a fluency of ideas.

The author's hypothesis that number ability would be found as a factor in much of the intermediate algebra is not substantiated by the loadings of Factor *D* in the algebra tests.

Factor E—Adaptability to a New Task

Test no.	Reference Tests	Study A Factor Loading	Study B Factor Loading
23	Similes	.30	.25
25	Letter-Blank Sentences	.44	.43
27	Block Counting	.37	.44
28	Letter Series	.30	.31
30	Family Trees	.32	.29
32	Similar Figures	.30	.35
34	Word-Word	.39	.33

All the tests in the study which involved new tasks for the subject had loadings above .30 on Factor *E*. None of the algebra tests had average loadings on Factor *E* above .14 for the two studies. Similarly, all those reference tests which involved tasks that the subjects had previously experienced had small loadings on this factor. Factor *E*, then, seems to represent the ease with which the student can learn a new task. It is noteworthy that this factor clearly appeared even in connection with such relatively simple new tasks as those included in this test battery. To some extent, Factor *E* is a factor of speed, but not the motor speed involved in performing simple manual tasks.

Multiple Choice versus Free Answer

Test 15, a test of converting statements to algebraic symbols, was constructed with multiple-choice alternatives. Test 16 was constructed with problems very similar to those of Test 15, but having no alternatives provided. The subject was required to give free answers. Similarly, Test 17 is a test of word problems, or statement problems, with multiple choice answers, whereas Test 18, with the same kind of problems, requires free answers.

TABLE 18

Comparison of Factor Loadings for Multiple Choice and Free Answer Tests

		Statements to Symbols		Word Problems	
		Multiple choice	Free answer	Multiple choice	Free answer
Factor	Study	Test no. 15	Test no. 16	Test no. 17	Test no. 18
A	A	.50	.42	.38	.34
Verbal Comprehension	B	.46	.40	.43	.46
B	A	.26	.28	.24	.40
Deductive Reasoning	B	.27	.26	.26	.32
C	A	-.02	.02	.12	-.06
Algebraic Manipulative Skill	B	-.15	.00	-.15	.10
D	A	.06	-.08	.14	.03
Number Ability	B	.10	.01	.05	.02
E	A	.09	.10	.13	.04
Adaptability to a New Task	B	.15	.14	.14	.06
Means	A	17.8	16.5	12.9	10.0
	B	19.6	18.0	14.6	11.3
Standard deviations	A	4.0	5.1	4.1	4.5
	B	2.9	3.5	3.9	4.8

The mean score of each multiple choice test for both Study A and Study B is higher than the mean score for the corresponding free answer test; however, the standard deviation of each multiple choice test is lower than that for the corresponding free answer test. The factor loadings on these four tests, Tests 15–18, as shown in Table 18, suggest little difference between the abilities needed to solve the multiple choice tests and those needed for solving the free answer tests. A study of the correlations of these two pairs of tests with the other 34 tests in the study, shown in Table 1, reveals this close similarity. The evidence tends to confirm the hypothesis: statement problems with multiple choice alternatives can be constructed to measure the same abilities as similar statement problems designed for free answers.

Questions Remaining to be Answered

This study would suggest the question: what serious consideration can be given to the factor loadings in a single factor analysis, by the Thurstone multiple-factor technique, if in two studies almost identical in nature only five of the twelve factors are congruent? It is true that on three of the twelve dimensions the overlap tests had only small projections. It is true, too, that if more severe criteria, as recommended by some writers, had been used to determine the completeness of the factor extraction, the number of factors in each study might have been as few as 6 or 7. Would it be wise in the future to conduct all factor analysis studies in two parts, to analyze each part separately, and then to test the two parts for congruence?

It is possible, on the other hand, that Tucker's technique crams a maximum amount of congruence into the first few factors, and thereby creates factors that are psychologically complex and somewhat obscure. This has added to the difficulty of identifying the factors with any large degree of assurance, and, too, it has added to the resulting high coefficients of correlation between the factors after rotation.

If a Study C has been made for a third successive year, similar in nature to Studies A and B, and if this analysis had been synthesized with the synthesis of the first two years, would the number of congruent factors have been diminished below five?

Factors of inductive reasoning, rote memory, spatial visualization, and fluency of expression—all expected by the author to appear in an analysis of intermediate algebra—failed to appear in the congruent factor space, unless they were undetected as a part of the five somewhat complex congruent factors.

No attempt was made to rotate the noncongruent factors to a meaningful interpretation, and therefore no effort was made to analyze or to interpret the seven factors of Study A and the seven factors of Study B that did not fall in the space congruent to the two studies. The interpretation of these factors would seem to have value only in the analysis of the study for each year separately and not for the over-all two-year study.

REFERENCES

[1] Army Air Force Aviation Psychology Program Research Reports. Printed classification tests. Report No. 5, 1947, 113–122.

[2] Coombs, C. H. A factorial study of number ability. *Psychometrika*, 1941, **6**, 161–189.

[3] French, J. W. The West Point tryout of the guidance battery, Part 2. Princeton: Educ. Test. Serv. Res. Bull. 55-6, 1955.

[4] Garrett, H. E. Differential mental traits. *Psychol. Rec.*, 1938, **2**, 259–298.

[5] Goodman, C. H. A factorial analysis of Thurstone's sixteen primary mental abilities tests. *Psychometrika*, 1943, **8**, 141–151.

[6] Harrell, W. A factor analysis of mechanical ability tests. *Psychometrika*, 1940, **5**, 17–34.

[7] Michael, W. B. Factor analyses of tests and criteria: a comparative study of two AAF populations. *Psychol. Monogr.*, 1949, **63**, No. 3.

[8] Peterson, D. A. Factor analysis of the new United States Navy basic classification test battery. Washington: U. S. Dept. of Commerce, 1946.

[9] Sisk, H. L. A multiple factor analysis of mental abilities in the freshman engineering curriculum. *J. Psychol.*, 1940, **9**, 165–177.

[10] Taylor, C. W. A factorial study of fluency in writing. *Psychometrika*, 1947, **13**, 239–262.

[11] Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. Chicago Press, 1947.

[12] Thurstone, L. L. Primary mental abilities. *Psychometric Monogr.*, No. 1. Chicago: Univ. Chicago Press, 1938.

[13] Thurstone, L. L. and Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monogr.*, No. 2. Chicago: Univ. Chicago Press, 1941.

[14] Tucker, L. R. A method for synthesis of factor analysis studies. Dept. of the Army, A.G.O. Personnel Res. Sec., Rep. No. 984, 1951.

Manuscript received 12/12/57
Revised manuscript received 2/21/59

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

REDUNDANCY IN TASK ASSIGNMENTS AND GROUP PERFORMANCE*

ROBERT B. ZAJONC AND WILLIAM H. SMOKE

UNIVERSITY OF MICHIGAN

The problem of combining abilities of group members to maximize the performance of the group as a whole is examined in terms of redundancy in task assignments. In particular, ways of distributing a given number of items of information among a given number of individuals to obtain the maximum probability of each item being recalled by at least one individual are studied. It is shown that there exists an optimal distribution scheme which is independent of the amount of material originally given, the size of the group, and individual differences in ability.

The problem of assessing determinants of group performance has been investigated from two different points of view. For the most part, studies in this area have been concerned with the effects of group variables, such as the presence of others [1, 2], cohesiveness [e.g., 8], leadership style [e.g., 5], and the like, on the performance of groups and of individuals working in groups. Recently, some attempts have been made to analyze the group product by means of a combinatorial analysis of individual abilities. Lorge and Solomon [4] performed such an analysis in the area of group problem solving, and Hays and Bush [3] have used it in group learning. In principle, this latter approach is analogous to that of Moore and Shannon in what they called the "crummy relay problem" [6], which refers to constructing reliable circuits out of unreliable relays. von Neumann [9] has shown that by using a number of components of limited unreliability a reliable machine may be constructed. Moore and Shannon have demonstrated that a reliable circuit may be designed by using arbitrarily unreliable relays. The increase in circuit reliability is obtained essentially by increasing the redundancy among relays. It would seem that the study of group performance in terms of redundancy among the abilities of individual members would hold considerable promise.

Consider, for example, a group of N individuals. Let H items of information be given to these individuals. The object is to recover this information from the group as a whole after some interval of time. For the present purposes it is irrelevant which individual remembers a particular item, although the item should be remembered by somebody in the group. There is evidence in

---

*This work was done under the sponsorship of the Behavioral Sciences Division, Air Force Office of Scientific Research, Contract AF 49(630)-33.

the area of individual recall to the effect that the proportion of items recalled is inversely related to the number of items originally assigned [10]. Consequently, the probability that a given item is remembered by a given individual is some inverse function of the number of items he was asked to learn. On the other hand, the probability that at least one individual of those assigned the given item remembers it increases with the number of individuals assigned the item. The first consideration implies minimizing the number of items per individual, the second maximizing it. The problem then is to discover the optimal distribution of items among individuals. The group as a whole is considered to remember an item when that item is remembered by at least one individual.

Case I

The following conditions are imposed on Case I.

(a) The probability $p(i, j)$ that item i is remembered by individual j is equal to the constant p, $0 \leq p \leq 1$, or to zero according to whether or not item i is assigned to j.

(b) Each individual is assigned the same number, h, of items. Thus $\sum_i p(i, j) = hp$.

(c) Each item is assigned to an equal number, n, of individuals. Thus the probability, P, that a given item is recalled by at least one individual is given by

$$P = 1 - \prod_n (1 - p) = 1 - (1 - p)^n, \tag{1}$$

the same for all items.

Under the conditions of Case I the problem is reduced to finding the assignment of items which generates the greatest value of P, i.e., finding values of p and n that will maximize P. If it is assumed that one is dealing not with number of items (a discrete measure) but with amount of material or amount of information (continuous measures), P may be regarded as a differentiable function of p. The necessary condition for P to be maximum as a function of p is that $dP/dp = 0$. Since h is a differentiable function of p, and since by (b) and (c) $Nh = Hn$, it follows that n is also a differentiable function of p. Thus, from (1)

$$\frac{dP}{dp} = 0 = -(1 - p)^n \left[\frac{dn}{dp} \log (1 - p) - \frac{n}{1 - p} \right]. \tag{2}$$

Given $p \neq 1$, $dP/dp = 0$ only if the bracketed expression is zero, or

$$\frac{dn}{dp} \log (1 - p) - \frac{n}{1 - p} = 0. \tag{3}$$

Since $n = Nh/H$,

$$\frac{dh}{dp}\log(1-p)-\frac{h}{1-p}=0. \tag{4}$$

Thus far the function relating p to h has not been specified. However, the condition in (4) holds for *any* set of values p and h which satisfy the relation $p = 1 - e^{k/h}$, for an arbitrary constant k. Thus

$$P = 1 - [1 - (1 - e^{k/h})]^{Nh/H} = 1 - e^{Nk/H}$$

is a constant since all the terms in the expression are constants. Thus, if the probability of a given item being recalled by a given individual were given by $p = 1 - e^{k/h}$, all assignments would be equally good. Under these conditions what is lost in p by assigning more items to each individual is gained by increasing n, and consequently the probability that an item is remembered by the group as a whole is independent of the assignment of items to the group members.

If, however, the relation between p and h is not given by $p = 1 - e^{k/h}$ but by some other function $p = f(h)$, then in general not all assignments will be equally good. In fact, on the basis of empirical data available in this area [10] it would appear that the function is of the form $p = e^{-k^2h^2}$, for an empirical parameter k which depends on such factors as time, nature of the material, its organization, meaningfulness, or the like. This function fits data gathered by Oberly [7] with $k = .10$. Assuming $p = e^{-k^2h^2}$ and finding dh/dp,

$$(1-p)\log(1-p) - 2p\log p = 0, \tag{5}$$

which is satisfied approximately for $p = .84$. Figure 1 shows the relation between P and p for selected values of k with $N/H = .01$.

Solving for h, $h = \sqrt{\log .84/-k^2} = .42/k$. Thus, the best assignment results when each individual is assigned $.42/k$ items. In terms of p, the maximum value of P obtains when each individual is given the number of items which would result in his forgetting about 16% of the material. Since the relationship between h, which denotes individual loads, and n, which reflects the amount of task-assignment redundancy, is known, the optimal amount of redundancy may be obtained. Thus, for Case I, $n = (.42/k)(N/H)$, the optimal amount of task-assignment redundancy.

It is rather interesting to note that the result obtained is entirely independent of the size of the group, N, and the number of items, H. Thus for any given number of individuals and any given number of items, $.42/k$ items per person represents the optimal assignment. Of course, the optimal amount of redundancy and consequently the maximum value of P vary with the ratio of individuals to items. The larger this ratio the higher the maximum possible value of P. Figure 2 represents the relationship between the maximum values of P and the N/H ratio for some selected values of the parameter k.

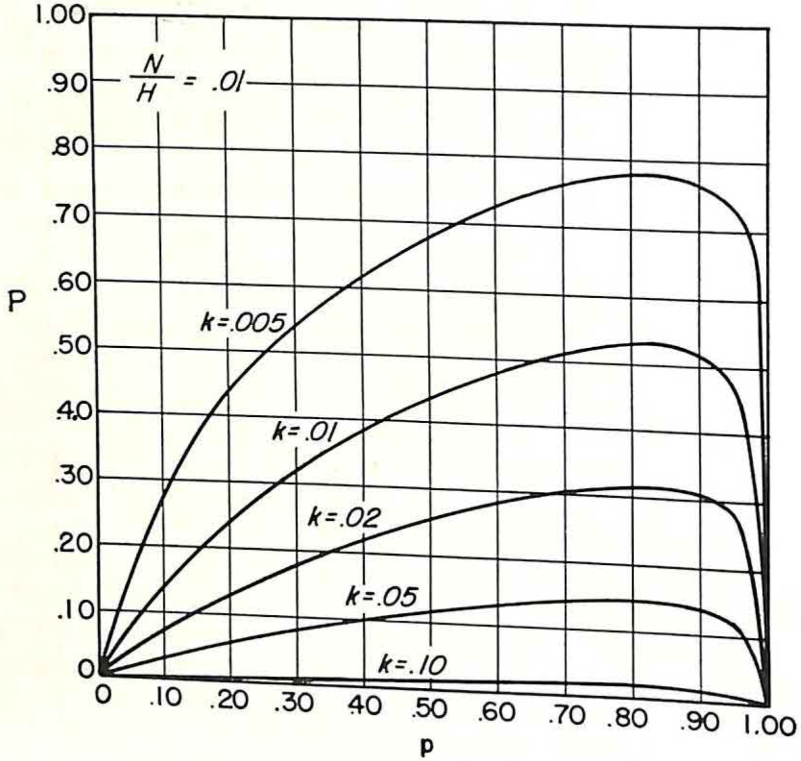


FIGURE 1

The Relationship Between P and p for Different Values of the Constant k.

The above solution was obtained by assuming h to be a continuous variable. Given a group of N individuals and a collection of H discrete items, it will not be possible in general to assign the items to individuals in more than a small number of ways. Hence the assignments do not vary continuously. As a matter of fact, for given values of N and H only some of all the possible assignments satisfy the conditions (b) and (c). It can be demonstrated that if D is the greatest common divisor of N and H, the number of assignments of H items to N group members satisfying (b) and (c) is equal to D, given that two assignments are not considered distinct when they assign the same number of items to each individual.

Case II

In Case II items of equal difficulties and individuals with equal recall capacities were considered. Now the case where there exist individual differences in recall will be examined.

Again a set of conditions is imposed.

(a) The probability, $p(i, j)$, that item i is remembered by individual j, is

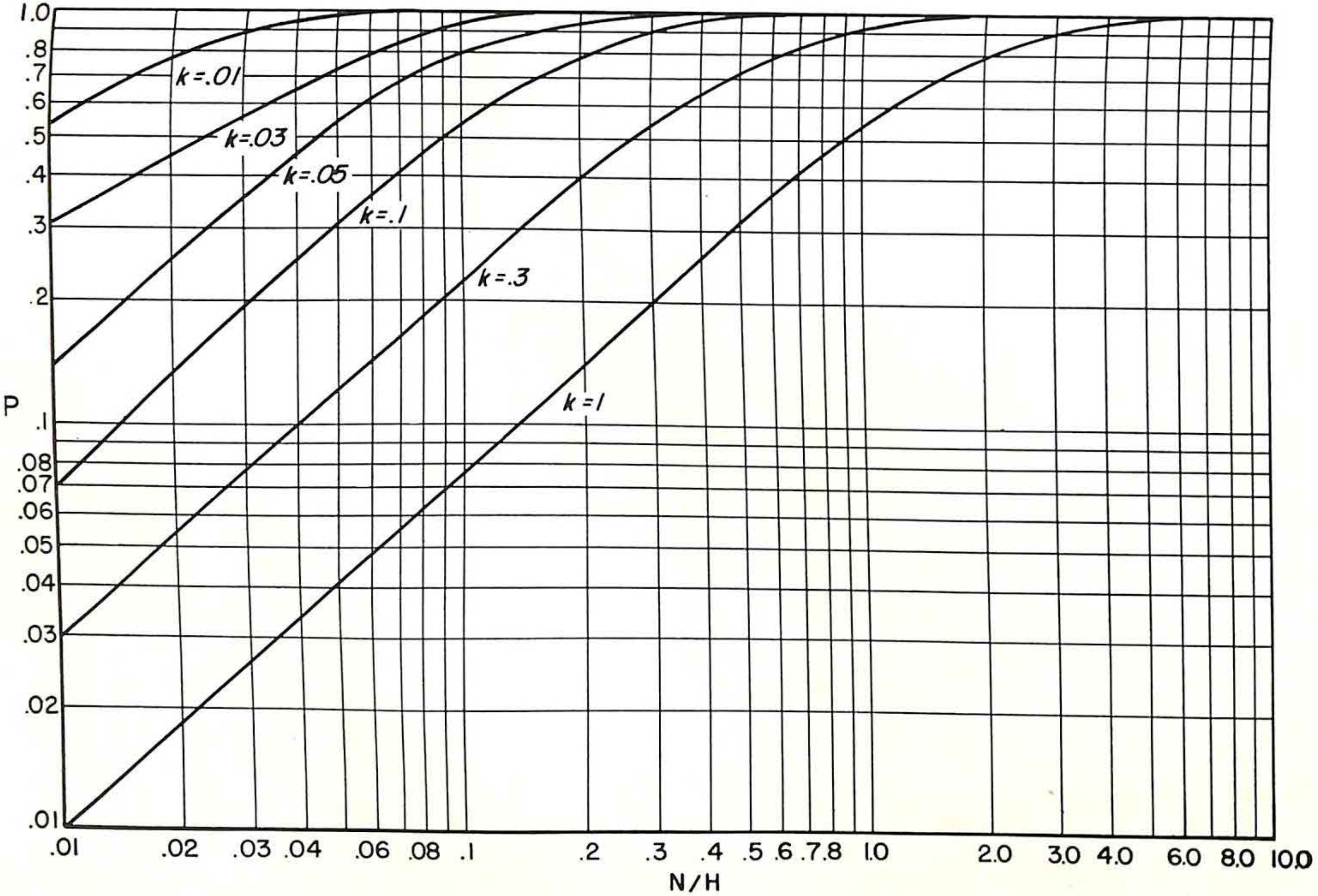


FIGURE 2

Maximum Values of P as a Function of N/H.

equal to the constant p, $0 \leq p \leq 1$, or to zero according to whether or not item i is assigned to j.

(b) The number, h_j, of items assigned to individual j is such that $p = e^{-k_j^2 h_j^2}$, where k_j is an empirical parameter obtained with respect to the individual j.

(c) The items are so distributed that for each item i the probability that i is remembered by at least one individual is equal to the constant $P' = 1 - \prod_j [1 - p(i, j)]$, the same for all items.

Thus, the above conditions imply that different individuals will be assigned different numbers of items, depending on their individual abilities to remember them. The conditions also imply that each item i is assigned to the same number of individuals n, or that the redundancy is equal for all items. Hence

$$P' = 1 - \prod_j [1 - p(i, j)] = 1 - (1 - p)^n. \tag{6}$$

Again, the necessary condition that P' be a maximum as a function of p is that $dP'/dp = 0$. Thus

$$\frac{dP'}{dp} = 0 = -(1 - p)^n\left[\frac{dn}{dp} \log (1 - p) - \frac{n}{1 - p}\right]. \tag{7}$$

If $p \neq 1$, then

$$(1 - p) \log (1 - p) \frac{dn}{dp} - n = 0. \tag{8}$$

In terms of the above conditions the number of assignments is equal to nH and to $\sum_j h_j$. Hence $nH = \sum_j h_j$. Thus

$$\frac{dn}{dp} = \frac{1}{H} \sum_j \frac{dh_j}{dp}.$$

From restriction (b),

$$\frac{dp}{dp} = 1 = -2k_j^2 h_j \frac{dh_j}{dp} e^{-k_j^2 h_j^2} = -2k_j^2 h_j \frac{dh_j}{dp} p = (2p \log p) \frac{1}{h_j} \frac{dh_j}{dp}.$$

Thus

$$h_j = (2p \log p) \frac{dh_j}{dp}$$

and

$$\sum_j h_j = (2p \log p) \sum_j \frac{dh_j}{dp}.$$

Therefore

$$\frac{dn}{dp} = \frac{1}{H} \sum_i \frac{dh_i}{dp} = \frac{1}{2p \log p} \left(\frac{1}{H}\right) \sum_i h_i = \frac{n}{2p \log p}$$

and

$$(1 - p) \log (1 - p) \frac{dn}{dp} - n = 1 - p) \log (1 - p) \frac{n}{2p \log p} - n,$$

or

$$(1 - p) \log (1 - p) - 2p \log p = 0. \tag{9}$$

Note that individual differences do not influence the solution, as (9) is satisfied for $p = .84$. However, the number of items, h_i , to be assigned to the different individuals will depend on their recall abilities which are reflected in the constants k_i . If the assignment of items to individuals satisfies (a) and (b), then for two individuals j_1 and j_2

$$\exp (-k_{j_1}^2 h_{j_1}^2) = p = \exp (-k_{j_2}^2 h_{j_2}^2)$$

or

$$k_{j_1} h_{j_1} = k_{j_2} h_{j_2} .$$

Thus

$$h_{j_2} = \frac{k_{j_1}}{k_{j_2}} h_{j_1}$$

and in general

$$h_{j_a} = \frac{k_{j_1}}{k_{j_a}} h_{j_1}$$

for each member j_a of the group. Since $\sum_j h_j = nH$,

$$nH = \sum_j \frac{k_{j_1}}{k_{j_a}} h_{j_a} = h_{j_1} k_{j_1} \sum_j \frac{1}{k_j}.$$

Hence

$$h_{j_1} = \left(\frac{nH}{k_{j_1}}\right) \left[\frac{1}{\sum_j \frac{1}{k_j}}\right],$$

and in general

$$h_{j_a} = \left(\frac{nH}{k_{j_a}}\right) \left[\frac{1}{\sum_j \frac{1}{k_j}}\right] \tag{10}$$

for each individual j. In this case individual differences are exploited by assigning fewer items to less able members and more items to the capable individuals.

While the value $p = .84$ is optimal under the restrictions specified above, it remains to be determined whether a different solution is obtained by relaxing restriction (a) such that $p(i, j)$, is no longer required to be a constant across individuals.

Conclusions

The solutions presented provide a standard against which empirical results may be compared. Empirical tests must conform to the predictions, otherwise the restrictions imposed on the solutions could not have been met. Thus by careful experimental controls it can be discovered what variables determine the departures from the prediction.

For instance, the conditions imposed above require that $p(i, j)$, the probability of the individual j recalling the item i, be constant. This, of course, necessitates a complete independence of the recall probabilities in terms of the items, as well as in terms of individuals. The probability $p(i_1, j_1)$ must be independent of $p(i_1, j_2)$ and of $p(i_2, j_1)$. Therefore, the cases examined are valid not for groups of interacting members but for collections of individuals working independently of one another.

This requirement, however, is not at all a shortcoming. On the contrary, it allows one to study the effects of group interaction on individual and group performance in recall. The existence of group interaction would probably lower the value of the parameter k, and this effect can be evaluated readily by empirical tests. The difference between the values of the parameter k for individuals working together and for individuals working alone would provide information on the effect of group interaction on individual performance, and the difference between the corresponding values of P, information concerning the effect of group interaction on group performance.

It is noted that the model presented is not restricted to recall, and that it may, with slight modifications, be applied to other behaviors, such as learning, problem solving, or decision making.

REFERENCES

[1] Allport, F. H. The influence of the group upon association and thought. *J. exp. Psychol.*, 1920, **3**, 152–182.

[2] Dashiell, J. F. An experimental analysis of some group effects. *J. abnorm. soc. Psychol.*, 1930, **25**, 190–199.

[3] Hays, D. G. and Bush, R. R. A study of group action. *Amer. sociol. Rev.*, 1954, **19**, 693–701.

[4] Lorge, I. and Solomon, H. Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*, 1955, **20**, 139–148.

[5] Maier, N. R. F. The quality of group decisions as influenced by the discussion leader. *Hum. Relat.*, 1950, **3**, 155–174.

[6] Moore, E. F. and Shannon, C. E. Reliable circuits using less reliable relays. Part I. *J. Franklin Inst.*, 1956, **262**, 191–208.
[7] Oberly, H. S. A comparison of the spans of "attention" and memory. *Amer. J. Psychol.*, 1928, **40**, 295–302.
[8] Schachter, S., Ellertson, N., and Gregory, D. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, **4**, 229–238.
[9] von Neumann, J. *Probabilistic logic*. Pasadena: California Inst. Technol., 1952.
[10] Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Manuscript received 1/22/59

Revised manuscript received 3/21/59

PSYCHOMETRIKA—VOL. 24, NO. 4
DECEMBER, 1959

A NOTE ON THE USE OF MOOD'S LIKELIHOOD RATIO TEST FOR ITEM ANALYSES INVOLVING 2 × 2 TABLES WITH SMALL SAMPLES

IRVING B. WEINER

UNIVERSITY OF MICHIGAN*

Mood's likelihood ratio test is generally considered an unreliable χ^2 approximation in 2 × 2 contingency tables containing expected cell frequencies less than five. Probability values were computed for 60 such tables as part of an item analysis for two 30-item alternate forms of a measure. The rank orders of the items, from best to worst differentiators, as determined separately by Mood's test and by Fisher's exact test correlated .97 for one form and .96 for the other.

Item analyses are often carried out not to determine the correlations of dichotomously scored items with a total test score, but rather to select a given number of items which best differentiates between high and low scorers. This type of analysis is facilitated by the use of 2 × 2 contingency tables which compare for each item the score (one or zero) of extreme scorers on the total test. The writer recently administered to groups of 25 and 22 subjects two 30-item alternate forms of a measure which he wanted to condense into a single 30-item form. Because of the restricted sampling, many of the expected cell frequencies in the resulting contingency tables were less than five. Since χ^2 approximations are generally not reliable with small cell frequencies [1], the probability values for the tables were computed directly by Fisher's exact test, and the 30 best items were identified accordingly. The probability values covered a wide range, from .008 to 1.00.

Probability values for the tables were also estimated by Mood's likelihood ratio test ([2], pp. 257–281), which is distributed approximately as χ^2 with $(r - 1)(s - 1)$ degrees of freedom for large samples. Mood states, however, that this large-sample approximation cannot be used without appreciable error when r and s both equal two. To evaluate this possibility with the 2 × 2 tables involved here, a comparison was made between the rank order of the items (from best to worst) for each of the two forms as determined by Mood's test and as indicated by the exact test. For one form the rank-order correlation (Spearman's rho) achieved between these two methods was .97 and for the other it was .96. Thus, although assumptions for its use were not met, Mood's test gave a very good indication of the *relative* probability values for the items. This may indicate that where ranking is the goal, as in the item analysis described, the χ^2 approximation by Mood's likelihood ratio test is an adequate statistical tool even with small samples.

---

*Now at University of Rochester School of Medicine.

REFERENCES

[1] Cochran, W. G. Some methods of strengthening the common χ^2 tests. *Biometrics*, 1954, **10**, 417–451.

[2] Mood, A. F. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1954.

Manuscript received 6/6/59

PUBLICATIONS RECEIVED FOR REVIEW

Ahmavaara, Y. *On the unified factor theory of mind.* Helsinki: Akademia Scientarum Fennica, 1957. Pp. 176.

Ahmavaara, Y. *The unified factor model.* Helsinki: Finnish Foundation for Alcohol Studies, 1958. Pp. 187.

Anderson, C. C. Function fluctuations. *British J. Psychol.*, Monograph Supplements XXX, 1958. Pp. 104.

Anderson, T. W. *Introduction to multivariate statistical analysis.* New York: Wiley, 1958. Pp. 374.

Anastasi, Anne. *Differential psychology.* New York: Macmillan, 1958. Pp. 664.

Ashby, W. R. *An introduction to cybernetics.* New York: Wiley, 1956. Pp. 295.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley, Calif.: Univ. California Press, 1956. Pp. 154.

Chapanis, A. *Research techniques in human engineering.* Baltimore: Johns Hopkins Press, 1959. Pp. 316.

Chernoff, H. and Moses, L. E. *Elementary decision theory.* New York: Wiley, 1959. Pp. 363.

Cornell, F. G. *Essentials of educational statistics.* New York: Wiley, 1956. Pp. 375.

Crow, E. L. et al. *Statistics manual.* China Lake, Calif.: U. S. Naval Ordnance Test Station. 1955. Pp. 288.

Dixon, W. J. and Massey, F. L. *Introduction to statistical analysis* (2nd Edition). New York: McGraw-Hill, 1956. Pp. 312.

Feller, W. *An introduction to probability theory and its applications.* New York: Wiley, 1957. Pp. 461.

Garrett, H. E. *Statistics in psychology and education.* New York: Longmans Green, 1958. Pp. 471.

Gross, N. et al. *Explorations in role analysis.* New York: Wiley, 1958. Pp. 379.

Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1956. Pp. 565.

Hakkinen, S. *Traffic accidents and driver characteristics.* Helsinki: Institute of Technology, Scientific Researches No. 13, 1958. Pp. 198.

Hohn, F. E. *Elementary matrix algebra.* New York: Macmillan, 1959. Pp. 305.

Kemeny, J. G. et al. *Introduction to finite mathematics.* New York: Prentice-Hall, 1957. Pp. 372.

Layton W. L. *Counseling use of the Strong Vocational Interest Blank.* Minnesota Studies in Student Personnel Work No. 8, Minneapolis: Univ. Minnesota Press, 1958. Pp. 40.

Loveday, R. *Statistics, a first course.* Cambridge: Cambridge Univ. Press, 1958. Pp. 121.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959. Pp. 149.

McCloskey, J. and Coppinger, J. M. (Eds.) *Operations research for management,* Vol. II. Baltimore: Johns Hopkins Press, 1956. Pp. 599.

McEwen, P. Figural after-effects. *British J. Psychol.*, Monograph Supplements XXXI, 1958. Pp. 106.

Neyman, J. (Ed.) *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: Univ. California Press, 1956. Pp. 184.

Odell, C. W. *A first course in educational statistics.* Dubuque, Ia.: Wm. C. Brown Co., 1957. Pp. 172.

Roy, S. N. *Some aspects of multivariate analysis.* New York: Wiley, 1957. Pp. 214.

Schachter, S. *The psychology of affiliation.* Stanford: Stanford Univ. Press, 1959. Pp. 141

Schreier, F. T. *Human motivation.* Glencoe, Ill.: Free Press, 1957. Pp. 256.

Senders, Virginia. *Measurement and statistics.* New York: Oxford Univ. Press, 1958. Pp. 598.

Simon, H. A. *Models of man.* New York: Wiley, 1957. Pp. 287.

Stephan, F. F. and McCarthy, P. J. *Sampling opinions.* New York: Wiley, 1958. Pp. 451.

Stone, Carol L. Percentages for integers 1 to 399. Stations Circular 341. Inst. Agri. Sciences, State College Washington, 1958. Pp. 399.

Sutcliffe, J. P. Task variability and the level of aspiration. *Australian J. Psychol.*, Monograph Supplement No. 2, 1955. Pp. 85.

Tate, M. W. and Clelland, R. C. *Nonparametric and shortcut statistics.* Danville, Ill.: Interstate Printers and Publishers, 1957. Pp. 171.

Terman, L. M. and Oden, Melita H. *The gifted group at mid-life.* Stanford: Stanford Univ. Press, 1959. Pp. 187.

Thrall, R. M. and Tornheim, L. *Vector spaces and matrices.* New York: Wiley, 1957. Pp. 318.

Thurstone, L. L. *The measurement of values.* Chicago: Univ. Chicago Press, 1959. Pp. 322.

Traxler, A. E. *Techniques of guidance.* New York: Harper, 1957. Pp. 314.

Viitamaki, R. O. *Personality traits between puberty and adolescence.* Helsinki: Finnish Academy of Science and Letters, 1956. Pp. 183.

BOOK REVIEWS

Jane Loevinger. Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, Monograph Supplement 9, 1957, 636–694.

A brief journal review cannot possibly do justice to a 60-page monograph which is so tightly packed with highly significant ideas and original points of view. The mission of the monograph is to present the point of view that the time has come to dispose of classical concepts of validity and to replace them with a concept in keeping with modern science. In particular, the proposal is that such concepts as predictive, concurrent, and content validities be swept aside as mere handmaidens of a psychotechnology of dubious scientific status. The clean sweep would also include such off-brand varieties of validity as factorial validity. In their place, it is proposed, the only concept of validity which be retained is that of construct validity, a concept which is developed more fully in this monograph than it has been developed elsewhere. It is argued that predictive, concurrent, and content validity are ad hoc concepts which have little to do with what the scientist does or what he needs to know.

The concepts of predictive and concurrent validity represent a concept of science which was typical of the late nineteenth century and which was presented by Karl Pearson in his famous *Grammar of Science*. A great amount of psychometric work today would represent an approach to the discovery of knowledge compatible with that presented by Pearson, and yet, in the hands of philosophers and logicians Pearson's approach has not fared well. In a sense, the introduction of the term construct validity by the APA Committee on Psychological Tests represents a major break with the classical correlational concepts of validity and an indication that psychometrics is recognizing the fundamental changes which have taken place in our conception of science since the days of Pearson. The central purpose of the monograph is to clarify the concept of construct validity and to begin to establish it as an idea of central importance in the development of a science of behavior.

Jane Loevinger is attempting to initiate a revolution in thinking in the area of psychometrics so that the instruments that evolve will be instruments for scientific advance rather than gadgets in a technology. Objective tests, she believes, are to play a central role in the development of psychological theory. Tests must be based on a theory of test behavior which is to be related to a theory of behavior in nontest situations. The elements in a psychological theory, it appears, are variables, but it is not clear whether these are or are not to be defined by objective tests.

The monograph proposes that the concept of validity be considered to include two components. One of these is the substantive component which is "the extent to which the content of the items included in (and excluded from?) the test can be accounted for in terms of the trait believed to be measured and the context of measurement. Context includes psychological theory and, in particular, 'the psychology of objective test behavior.' " The writer suspects that what Loevinger means by psychological theory may be different from what he means by psychological theory, but the concept of psychological theory involved in the document under review is not developed to the point where such a comparison is possible. Within the field of psychometrics, test items are often discussed as if their properties could be described in the same terms as other stimuli, such as a pure tone or a flash of light or an electrical discharge. The experimental psychologist is likely to consider the properties of test items as stimuli as the consequence of antecedent learning conditions and that the properties of the items must be measured in terms of those conditions. Thus a semantic count provides a rough measure of typical exposure to particular words used in particular contexts, and this measure in turn can be used to predict some of the responses that will occur to the particular item when it is presented as a part of a test. Such a measure of semantic frequency is one dimension along which the substantive component of a test item can be fixed. Presumably there would be other components too. But how would the substantive characteristics of items be determined in the case of items from personality tests? How can one tie down the antecedent conditions which generate the particular stimulus values of the items? In the Loevinger monograph such problems are not elaborated and the reviewer is left with the uneasy feeling that Loevinger is concerned with intuitively derived psychological theory rather than that in which the constructs are hypothetical and include unobservables. This is no real criticism of the monograph but only points up the long and difficult road that lies ahead of the person who embarks on a program of integrating psychometrics with psychological theory.

The second major component of validity proposed in the monograph is the structural component, which is a much better defined concept than is the substantive component. The structural component refers to "the extent to which structural relations between test items parallel the structural relations of other manifestations of the trait being measured." Interpreted in terms of the language of systematic psychology, one might say that structural validity implies that relations between responses within test situations should parallel in some way the relations discovered between responses in nontest situations. The structural component refers to the extent to which a test is a manifestation of a law relating responses to responses. Such laws, commonly described as R–R laws, have been generally considered to be rather low-level laws; that is to say, the uses which they have are limited, and the boundary conditions for them can rarely be specified. Herein lies one of the major differences between psychometrics and experimental psychology. In psychometrics, R–R laws are sought, while the experimental psychologist seeks S–R laws because of the particular advantages which they possess. A thoroughgoing integration of psychometrics and current psychological theory would require that psychometrics become a means of developing S-R laws, but perhaps it is much too early to hope for this development. Loevinger has made a beginning by identifying some of the problems which this integration involves.

The monograph provides an overview of various structural models that have been used and the problems which they involve. Under this heading the author has included quantitative models, class models, and dynamic models. The material could be dry to read, but it is not. Novel ways of looking at the problems involved add a continuous source of freshness to the discussion. Indeed, the incidental comments alone make this monograph worth reading, but it contains much more than incidental material.

Loevinger writes with a contagious enthusiasm for her subject. Whether the reader agrees or disagrees with the arguments presented, he will come away from the monograph with the feeling that he has lived through a refreshing and worthwhile experience.

University of Utah ROBERT M. W. TRAVERS

L'Analyse Factorielle et ses Applications. Paris: Centre National de la Recherche Scientifique, 1955. Pp. 431.

For five days in July, 1955, 37 factor analysts and scholars interested in factor analysis from 12 countries, met in Paris to discuss papers on factor theory and its applications. This book is the result of their meeting.

The conference was conceived by Henri Laugier, well known as a physiologist and as a scientist with a strong interest in applied psychology. It was supported, in part, by a grant from the Rockefeller Foundation. Listed home countries for those actually present were: France, 15; Great Britain, 5; Sweden, 3; Egypt, Germany, Israel, Switzerland, and the United States, 2 each; and Belgium, Canada, Spain, and Yugoslavia, 1 each.

Among those who had been invited but who were not actually able to be present were two who were represented by papers: Sir Cyril Burt and L. L. Thurstone. Their messages of regret are on early pages of the report. In addition to his formal paper, Thurstone contributed a resumé of the Conference on Factor Studies held at Educational Testing Service, Princeton, New Jersey, in November, 1951.

Papers were circulated in advance, and the meetings, with the help of translators, concentrated on discussion. Of 22 papers, 9 were originally in French, 13 in English. In the book all papers and summaries of the discussions are presented in French. The following list of papers is a good guide to the nature of its content.

"Current problems and new methods in factor analysis," L. L. Thurstone.

"The Uppsala Symposium on psychological factor analysis, Uppsala, March 17–19, 1953," E. A. Peel.

"Dimensions of intellect," J. P. Guilford.

"Factor analysis: methods and results," Sir Cyril Burt.

"Le problème général de la recherche et de la nature des facteurs en psycho-physiologie," (The general problem of research on factors and of their nature in psycho-physiology), H. Pieron.

"Relations of the newer multivariate statistical methods to factor analysis," H. Hotelling.

"Remarques sur l'analyse factorielle de Hotelling et comparison avec les méthodes centroides," (Remarks on Hotelling's factor analysis and a comparison with centroid methods), H. Pineau.

"A statistical test for the stability of simple structure," R. Bargmann.

"Psychological meaning of factor analysis as a research method," M. Yela.

"Les facteurs psychologiques: quelques remarques sur leur nombre, leur identification, leur nature," (Psychological factors: remarks on their number, identification, and nature), Miss G. Bernyer.

"Facteurs observés et facteurs théoriques en psychologie," (Observed and theoretical factors in psychology), M. Reuchlin.

"The radex approach to factor analysis," L. Guttman.

"Utilisation du schéma de Spearman, dans le calcul des images," (The use of a Spearman formulation in the calculation of images), J. M. Faverge.

"Factor analysis and the problem of validity," H. J. Eysenck.

"Emploi de l'analyse factorielle dans l'étude de la variabilité biologique," (The use of factor analysis in the study of biological variability), E. Schreider.

"Application de l'analyse factorielle à l'étude de la mortalité," (Application of factor analysis in the study of mortality), S. Ledermann.

"Observations theoriques sur l'analyse factorielle linéaire et générale," (Theoretical remarks on general, linear factor analysis), G. Darmois.

"Nouvelle méthode de statistique mathématique pour l'estimation des facteurs et de leur écart-type en analyse factorielle," (A new method in mathematical statistics for the estimation of factors and their standard deviations in factor analysis), P. Delaporte.

Discussion of a proposal of R. B. Cattell, "A universal index for psychological factors."

"Trends of research in space abilities," A. H. El Koussy.

"Factor analysis of achievement tests: methodological considerations and some empirical findings," T. Husen and S. Henrysson.

"The factor analysis of person correlations and the use of independent determiners to identify the factors," E. A. Peel.

"Rapport de synthèse," (Summary of the conference), M. Reuchlin.

These papers and related discussions depict remarkably clearly the contemporary state of the science and art of factor analysis. Contributors to the colloquium represented a broad spectrum of views, both on theoretical issues and on practical applications of factor methods.

There were differences in historical orientation. Thurstone's paper traced the origins of factor analysis to the contributions of Spearman, while Burt's paper ascribed the first version of the two-factor theorem to Galton and the method of principal axes to Pearson. Burt stated that the fundamental formula used by himself and his associates prior to 1925 was the "centroid equation," ultimately adopted by Thurstone as the basis of the centroid method.

Controversies centering around the two-factor approach of Spearman as contrasted with the multifactor approach of Thurstone were not in evidence. However, the theoretical position of Burt and his followers seemed to occupy middle ground as far as empirical findings were concerned. A general factor was still important in mental test material but there was generous provision for group factors and special abilities.

Considerable attention was given to mathematical issues, especially on questions relating to sampling. Hotelling pointed out that many of the original objectives of factor analysis (but not all) could be better reached by other multivariate methods. He treated factor analysis as a means of estimating the dimensionality of a population, pointing out that the greatest use of present factor procedures was to suggest hypotheses susceptible of being proved more objectively by other methods. Darmois, mathematical statistician at the Sorbonne, pointed out a number of the probability considerations arising in factor analytic practices, while Delaporte presented a method of estimating the standard error of a factor loading. In a somewhat parallel development, Bargmann's discussion of simple structure is of considerable interest.

Several of the papers, notably those of Thurstone, Guilford, Burt, and Hotelling, presented points of view that would have been quite familiar to readers of *Psychometrika* in 1955. Thurstone was pleased that multiple factor analysis had been judged sufficiently useful to justify an international conference, and indicated some of its eventual scientific possibilities. Guilford summarized his findings of the preceding six years, using the general framework of the centroid method and orthogonal rotations to simple structure, and seeking factors of psychological significance in high aptitude personnel, especially in the areas of reasoning and creative intelligence.

The volume is an important source of information on Guttman's radex approach to factor analysis. In 1955, Guttman's contributions to the classification of patterns of matrices of correlations were not yet well known. His views of the simplex and circumplex were elaborated not only in his paper but also in numerous discussions.

To this reviewer, one of the most surprising (and pleasing) papers was that of Pieron, who is well known to American psychologists, but hardly as a factor analyst. Yet such is Pieron's versatility that, on the basis of reading 150 papers, he presented a thoughtful resumé not only of factor theory but also of applications in areas of special interest to investigators of problems in physiological psychology.

The future historian of factor analysis will find this book a source of considerable information on trends and unsolved problems in the mid-fifties. The contemporary factor analyst will find some material not readily available elsewhere, as well as a number of stimulating papers and discussions.

Washington University PHILIP H. DUBOIS

MEMBERS OF THE PSYCHOMETRIC SOCIETY*

Robert P. Abelson, 272 Great Oak Road, Orange, Connecticut
Sidney Adams, 242 North Granada Street, Arlington, Virginia
C. J. Adcock, Victoria University College, Wellington, New Zealand
Marvin Adelson, Hughes Aircraft Company, Building 384, Mail Station 1066, P. O. Box 2097, Fullerton, California
Dorothy C. Adkins, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina
*Joel W. Ager, Jr., 505 W. Ten Mile Road, Pleasant Ridge, Michigan
Yrjo Ahmavaara, Koroistentie 15 A8, Helsinki, Finland
Edward F. Alf, Jr., 2866 Monarch Street, San Diego 11, California
Roger B. Allison, Jr., Department of Psychology, 156-168 Westminster Avenue, Salt Lake City 15, Utah
Earl A. Alluisi, P. O. Box 235, Emory University, Atlanta 22, Georgia
*Vladimir V. Almendinger, Jr., 26 Beechcroft Street, Brighton 35, Massachusetts
Mrs. Carol H. Ammons, P. O. Box 1441, Missoula, Montana
Robert B. Ammons, P. O. Box 1441, Missoula, Montana
Wade S. Amstutz, R.F.D. No. 1, Yellow Springs, Ohio
Anne Anastasi, Graduate School, Fordham University, New York 58, New York
Oliver F. Anderhalter, Bureau of Institutional Research, 15 North Grand Boulevard, St. Louis 3, Missouri
*G. Ernest Anderson, Jr., Apartment 12-A, 36 Linnaean Street, Cambridge 38, Massachusetts
Nancy S. Anderson, Department of Psychology, University of Maryland, College Park, Maryland
Norman Anderson, University of California, Los Angeles 24, California
Richard J. Anderson, 1203 12 Avenue, N.W., Gainesville, Florida
T. W. Anderson, Department of Mathematical Statistics, Columbia University, New York 27, New York
T. G. Andrews, Department of Psychology, University of Maryland, College Park, Maryland
William H. Angoff, Educational Testing Service, Princeton, New Jersey
Valentine Appel, R.F.D. No. 2, Peekskill, New York
E. James Archer, Department of Psychology, University of Wisconsin, Madison, Wisconsin
*Richard F. Arnold, 1572G Spartan Village, East Lansing 2, Michigan
Wilhelm Arnold, Meyer-Olberslebenstrasse 7, Wurzburg, Germany
Philip Ash, 645 South Bruner Street, Hinsdale, Illinois
Richard C. Atkinson, Psychology Department, University of California, Los Angeles 24, California
E. H. Baade, Eiksveien 58, Røa, Oslo, Norway
Melany E. Baehr, 1151 East 85th Street, Chicago 19, Illinois
Andrew R. Baggaley, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee 11, Wisconsin
Donald E. Baier, 909 Westholm Road, Schenectady 9, New York
*Mrs. Joan Hauser Bailey, 13038 Premiere Avenue, Downey, California
Frank B. Baker, 6400 Hampshire Place, Minneapolis 21, Minnesota

*An asterisk preceding a name indicates student membership.

Paul J. Barnes, 80 Dumwoodie Street, Yonkers, New York
A. S. Barr, Professor of Education, University of Wisconsin, Madison 6, Wisconsin
*Joan A. Barr, 1136 Michigan Avenue, Ann Arbor, Michigan
*Alan R. Bass, Department of Psychology, University of Illinois, Urbana, Illinois
*Herman D. Bates, 1308 Hutchins Avenue, Ann Arbor, Michigan
Robert L. Batterton 4626 W. Pt. Loma Boulevard, San Diego, California
William F. Battig, Psychological Laboratory, University of Virginia, Charlottesville, Virginia
Nancy Bayley, Psychology Laboratory, Institute of Mental Health, National Institutes of Health, Bethesda 14, Maryland
David C. Beardslee, Psychological Laboratory, Wesleyan University, Middletown, Connecticut
Katharine Pease Beardsley, 3817 Van Ness Street, N.W., Washington 16, D.C.
Harold P. Bechtoldt, Department of Psychology, State University of Iowa, Iowa City, Iowa
Ralph C. Bedell, 3700 Massachusetts Avenue, N.W., Washington 16, D.C.
A. W. Bendig, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
George K. Bennett, Psychological Corporation, 304 East 45th Street, New York 17, New York
Purnell Benson, 21 Maple Avenue, Madison, New Jersey
*Carl E. Bereiter, 104-G Eagle Heights, Madison 5, Wisconsin
Marvin H. Berkeley, 6767 Southwest, St. Louis 17, Missouri
Rudolph G. Berkhouse, 437 Monticello Boulevard, Alexandria, Virginia
Mitchell M. Berkun, USALHRU, Box 787, Presidio of Monterey, California
Morton A. Bertin, Florida School for the Deaf and Blind, St. Augustine, Florida
*Robert Besco, Department of Psychology, Purdue University, Lafayette, Indiana
Rev. Stanley John Bezuszka, S. J., Chairman, Department of Mathematics, Boston College, Chestnut Hill 67, Massachusetts
Robert C. Bilger, 1420 Golden Avenue, Ann Arbor, Michigan
Arnold Binder, Department of Psychology, Indiana University, Bloomington, Indiana
Joseph D. Birch, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Allan Birnbaum, Department of Mathematical Statistics, Columbia University, New York 27, New York
James E. Birren, National Institute of Mental Health, Bethesda 14, Maryland
Donald B. Black, 8522-82 Avenue, Edmonton, Canada
Paul Blommers, State University of Iowa, Iowa City, Iowa
Frances Blumenthal, 6 Landscape Avenue, Yonkers 5, New York
R. Darrell Bock, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
Robert F. Boldt, AGTL-S, Department of Army, T.A.G.O., Washington 25, D.C.
Nicholas A. Bond, Jr., 4432 Don Miguel Drive, Los Angeles 8, California
Guido Borasio, Psychology Department, Washington University, St. Louis 5, Missouri
Edgar F. Borgotta, 34-18 73rd Street, Jackson Heights, New York
Harry Bornstein, 1308 S. George Mason Drive, Arlington 4, Virginia
Henry Borow, The General College, University of Minnesota, Minneapolis 14, Minnesota
Robert A. Bottenberg, 2666 West Summit, San Antonio, Texas

Rev. W. A. Botzum, C.S.C., University of Portland, Portland 3, Oregon
Lyle E. Bourne, Jr., Department of Psychology, University of Utah, Salt Lake City, Utah
Eugene A Bouvier, 3229 Ettrick Street, Los Angeles 27, California
James E. Boyce, 1543 Witherbee, Birmingham, Michigan
Henry R. Brenner, Raytheon Manufacturing Company, 1360 Soldiers Field Road, Brighton, Massachusetts
Francois Bresson, 133 rue La Marck, Paris 18, France
Claude F. Bridges, c/o Evaluation Center, Fort Benjamin Harrison, Indianapolis 49, Indiana
George E. Briggs, Laboratory of Aviation Psychology, The O.S.U. Research Center, 1314 Kinnear Road, Columbus 8, Ohio
Harold F. Bright, 405 Branch Road, S. G., Vienna, Virginia
Hubert E. Brogden, Personnel Research Branch, T.A.G.O., D.A., Washington 25, D.C.
F. Robert Brush, Department of Psychology, University of Maryland, College Park, Maryland
Glenn L. Bryan, Department of Psychology, Electronics Research, University of Southern California, 3516 University Avenue, Los Angeles 7, California
Joseph G. Bryan, 45 Strawberry Hill Court, Stamford, Connecticut
*Mark Philip Bryden, 3684 McTavish Street, Montreal 2, Quebec, Canada
Donald N. Buckner, 1728 Strand, Hermosa Beach, California
Hermann F. Buegel, University of North Dakota, Box G, Grand Forks, North Dakota
Eugene I. Burdock, 722 West 168th Street, New York 32, New York
George G. Burgess, R. D. No. 1, Box 265-6, Bellefonte, Pennsylvania
Laverne K. Burke, 2475 Virginia Avenue, N.W., Apt. 525, Washington 7, D.C.
Thomas H. Burkhalter, Q.M. Research and Engineering Command, U.S. Army, Natick, Massachusetts
Oscar K. Buros, 220 Montgomery Street, Highland Park, New Jersey
Raymond H. Burros, Technical Operations, Inc., Box 37, Fort Monroe, Virginia
Harold E. Burtt, Department of Psychology, Ohio State University, 1945 North High Street, Columbus 10, Ohio
Allen C. Busch, 4074 Ellis Avenue, Batavia, Ohio
John Caffrey, Director of Research, Palo Alto Schools, 25 Churchill Street, Palo Alto, California
Joel T. Campbell, 24700 Alberton Road, Euclid 23, Ohio
*Vincent N. Campbell, Department of Psychology, University of Colorado, Boulder, Colorado
Joan H. Cantor, Psychology Department, Peabody College, Nashville, Tennessee
Frederick O. Carleton, 931 Michigan Avenue, Apt. 2E, Evanston, Illinois
*Richard M. Carr, Tanglewood Lane Apartments, Apt. D-2, King of Prussia, Pennsylvania
John B. Carroll, 7 Kirkland Street, Cambridge 38, Massachusetts
Edwin Calvin Carterette, Department of Psychology, University of California, Los Angeles 24, California
Raymond B. Cattell, 106 S. Goodwin Street, Urbana, Illinois
Randall M. Chambers, Aviation Medical Acceleration Laboratory, USN Air Development Center, Johnsville, Pennsylvania
Robert E. Chandler, 13-264 General Motors Building, Detroit 2, Michigan
Jean Paulsen Chapman, 232 Conn Terrace, Lexington, Kentucky
Robert L. Chapman, 1801 Sea Drift Drive, Corona del Mar, California

Henry Chauncey, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
William B. Chew, Assistant Director, Employee Research Section, 13-264 General Motors Building, Detroit 2, Michigan
Ralph McC. Chinn, P. O. Box 591, Emory University, Georgia
Paul R. Christensen, 1551 Gates Avenue, Manhattan Beach, California
Richard R. Clampitt, Columbus Receiving Hospital, University Health Center, Columbus 10, Ohio
*Cherry Ann Clark, P. O. Box 304, South Pasadena, California
Kenneth E. Clark, Department of Psychology, University of Minnesota, Minneapolis, Minnesota
Stephen C. Clark, 1759 LaPaz Road, Altadena, California
W. V. Clemans, National Board of Medical Examiners, 133 S. 36th Street, Philadelphia 4, Pennsylvania
Dorothy M. Clendenen, The Psychological Corporation, 304 East 45th Street, New York 17, New York
Norman Cliff, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
*Bart B. Cobb, Jr., 534 Byrnes Drive, San Antonio, Texas
Herbert H. Coburn, 60 Commonwealth Avenue, Boston 16, Massachusetts
Bernard P. Cohen, 1325 Josephine Street, Berkeley 3, California
Jacob Cohen, 209 Concord Road, Yonkers, New York
Joseph W. Cole, 331 Brookview Drive, Rochester 17, New York
Andrew L. Comrey, Jr., Department of Psychology, University of California, Los Angeles 24, California
Herbert S. Conrad, Research and Statistical Service, U.S. Office of Education, Washington 25, D.C.
Desmond L. Cook, Department of Psychology, Purdue University, West Lafayette, Indiana
Ellsworth B. Cook, Naval Medical Field Research Laboratory, Camp Lejeune, North Carolina
William E. Cook, Alton State Hospital, Alton, Illinois
Clyde H. Coombs, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Herman A. Copeland, 909 Atwood Road, Chesney Downs, Philadelphia 18, Pennsylvania
Walter Coppock, Surgery Department, University of Oregon Medical School, Portland, Oregon
Mary Corcoran, 301 Burton Hall, University of Minnesota, Minneapolis 14, Minnesota
Francis G. Cornell, Engelhardt, Engelhardt, Leggett, and Cornell, 221 W. 57th Street, New York 19, New York
Don J. Cosgrove, 2967 Douglan Terrace, Cincinnati 13, Ohio
John Whealdon Cotton, Department of Psychology, Northwestern University, Evanston, Illinois
Sarah Counts, 2354 Hilgard Avenue, Berkeley 9, California
John T. Cowles, University of Pittsburgh, Pittsburgh 13, Pennsylvania
Gertrude M. Cox, North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina
Henry M. Cox, Director, Bureau of Instructional Research, New Administration Building, Room 104, University of Nebraska, Lincoln 8, Nebraska
Kenneth J. Cox, 35 W. Park Avenue, Hamilton, Ontario, Canada
Lolafaye Coyne, The Menninger Foundation, Topeka, Kansas
Elliot M. Cramer, Psychology Department, The Johns Hopkins University, Baltimore, Maryland
Clarke W. Crannell, 120 South Main Street, Oxford, Ohio

William J. E. Crissy, 51-01 39th Avenue, Sunnyside, Long Island City 4, New York
Lee J. Cronbach, 1007 S. Wright Street, Champaign, Illinois
Dorothea J. Crook, Psychology Department, Tufts University, Medford 55, Massachusetts
*Daniel B. Cruse, 1211-E Brackenridge Apts., Austin, Texas
E. E. Cureton, 1846 Prospect Place, S.E., Knoxville 15, Tennessee
Louise W. Cureton, 1846 Prospect Place, S.E., Knoxville 15, Tennessee
Manuel Cynamon, Office of Research and Testing, Brooklyn College, Brooklyn 10, New York
W. Grant Dahlstrom, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina
John T. Dailey, 3516 S. Wakefield, Arlington 6, Virginia
Karl M. Dallenbach, 2106 Meadowbrook Drive, Austin, Texas
Fernando Rodrigues da Silveira, Rua Francisca Sales 90 (Jacarepagua), Rio de Janeiro, Brazil
Melvin D. Davidoff, U.S. Public Health Service, Gerontology Branch, City Hospital, Baltimore 24, Maryland
Frederick B. Davis, Educational Clinic, Hunter College, 693 Park Avenue, New York 21, New York
Paul C. Davis, Los Angeles Pacific College, 625 Coleman Avenue, Los Angeles 42, California
Hugh M. Davison, 419 Martin Terrace, State College, Pennsylvania
Robert Dear, System Development Corporation, 2400 Colorado Avenue, Santa Monica, California
Walter L. Deemer, Jr., 3624 Appleton Street, N.W., Washington 8, D.C.
James W. Degan, The MITRE Corporation, P. O. Box 208, Lexington 73, Massachusetts
Adriaan DeGroot, Roemer Visscher straat 11, Amsterdam, Holland
Richard L. Deininger, Bell Telephone Laboratories, Murray Hill Laboratory, Murray Hill, New Jersey
John E. S. de Jung, 12408 Dalewood Drive, Silver Spring, Maryland
Dr. Francisco del Olmo, Industrial Relations Department, Creole Petroleum Corporation, Apartado 889, Caracas, Venezuela
Robert G. Demaree, 803 Meadowbrook Drive, North Tonowanda, New York
J. C. Denton, Psychological Business Research, 815 Superior Avenue, N.E., Cleveland 14, Ohio
Mrs. Lillian Capo de Rivero, Box 395, Roosevelt, Puerto Rico
Kern William Dickman, 7 Buena Vista Court, Urbana, Illinois
Harvey F. Dingman, Pacific State Hospital, Box 100, Pomona, California
Wilfrid J. Dixon, Department of Preventive Medicine, University of California Medical Center, Los Angeles, California
Jerome E. Doppelt, 18 Mansfield Road, White Plains, New York
William F. Dossett, 4535 Santo Tomas Drive, Los Angeles 8, California
*Howard J. Douglass, Department of Psychology, Rensselaer Polytechnic Institute, Troy, New York
Paul L. Dressel, Director of Evaluation Services, 5A Berkey Hall, Michigan State University, East Lansing, Michigan
Raymond J. Dry, Life Insurance Agency Management Association, 855 Asylum Avenue, Hartford 5, Connecticut
Philip H. DuBois, Department of Psychology, Washington University, St. Louis 5, Missouri
Edmund Emil Dudek, U. S. Naval Personnel Research Field Activity, San Diego 52, California
Robert D. Dugan, Director of Agency Research, State Farm Insurance Companies, 112 E. Washington Street, Bloomington, Illinois

Frank M. DuMas, Department of Psychology, Montana State University, Missoula, Montana
Jack W. Dunlap, Lockwood Road, Riverside, Connecticut
Theodore F. Dunn, 2375 East 3 Street, Brooklyn, New York
Paul S. Dwyer, 403 Lenawee Drive, Ann Arbor, Michigan
Henry Dyer, 153 Jefferson Road, Princeton, New Jersey
Robert L. Ebel, 20 Nassau Street, Princeton, New Jersey
Harold Edgerton, 4 Nursery Road, New Canaan, Connecticut
Allen L. Edwards, Department of Psychology, University of Washington, Seattle 5, Washington
Donald Ehrman, 460 Forest Avenue, Palo Alto, California
Ahmed El Senoussi, 13681 Sunset Boulevard, Pacific Palisades, California
Max D. Engelhart, 1648 E. 85th Place, Chicago 17, Illinois
*Jean Engler, Department of Psychology, University of London, Gower Street, W. C. 1, London, England
Sidney Epstein, 922 Oak Lane, Apt. 2, Menlo Park, California
Donald Estavan, System Development Corporation, 2500 Colorado, Santa Monica, California
Mrs. Katherine W. Estes, Department of Psychology, Indiana University, Bloomington, Indiana
Harold T. Fagin, 136 Jay Street, Johnson City, New York
Robert F. Fagot, Psychology Department, University of Oregon, Eugene, Oregon
Edmund W. Faison, Prudential Plaza, Chicago 8, Illinois
Chung-Teh Fan, 117 N. McKinley Avenue, Endicott, New York
Mary E. Farabaugh, 4201 Massachusetts Avenue, N.W., Washington 16, D.C.
S. David Farr, 165 Greentree Road, Tonowanda, New York
*Bruce D. Faulds, Eno Hall, Princeton University, Princeton, New Jersey
J. M. Faverge, 13 Rue Paul Chautard, Paris xv, France
Leonard S. Feldt, Iowa Testing Programs, East Hall, Iowa City, Iowa
George A. Ferguson, Department of Psychology, Peterson Hall, McGill University, 3684 McTavish Street, Montreal, Quebec, Canada
Leon Festinger, Department of Psychology, Stanford University, Stanford, California
Victor Fields, 812 Arrington Drive, Silver Spring, Maryland
Gordon Fifer, Apt. BB-324 West 88th Street, New York City 24, New York
Warren G. Findley, Assistant Superintendent for Pupil Personnel Services, Atlanta Board of Education, 10th floor, City Hall, Atlanta 3, Georgia
Clyde L. Fischer, 173 Duffant Street, Santurce, Puerto Rico
Joshua A. Fishman, Greenfield Center for Human Relations, University of Pennsylvania, Philadelphia 4, Pennsylvania
David J. Fitch, 4901 Niagara Road, College Park, Maryland
Robert Fitzpatrick, 4423 54th Avenue, S.W., Seattle 16, Washington
John C. Flanagan, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
*Ronald Lynn Flaugher, 3548 Grandin Road, Cincinnati 26, Ohio
Stewart E. Fliege, 3540 Woodcliff Road, Sherman Oaks, California
William J. Flynn, Jr., 21316 West Lake Road, Rocky River 16, Ohio
Frank B. Ford, 163 Green Haven, San Antonio, Texas
Raymond W. Frankman, Department of Psychology, University of Illinois, Urbana, Illinois
Norman Frederiksen, 163 Hamilton Avenue, Princeton, New Jersey
John W. French, Educational Testing Service, P. O. Box 592, Princeton, New Jersey
*Alan Frenkel, Hall of Residence, 550 First Avenue, New York, New York
*Morton P. Friedman, Psychology Department, Indiana University, Bloomington, Indiana

Benjamin Fruchter, 1216 Norwood Road, Austin, Texas
Edmund F. Fuchs, 1806 Florin Street, Silver Spring, Maryland
Marguerite S. Gadel, General Planning & Research Division, Prudential Insurance Company, Newark 1, New Jersey
Eugene Galanter, 3440 Walnut Street, Philadelphia 4, Pennsylvania
Henrietta L. Gallagher, 168 Nassau Street, Princeton, New Jersey
Paul A. Games, Department of Psychology, Ohio University, Athens, Ohio
Eric F. Gardner, Director, Division of Advanced Studies, School of Education, 200 Slocum Hall, Syracuse 10, New York
Wendell R. Garner, Department of Psychology, Johns Hopkins University, Baltimore 18, Maryland
Henry E. Garrett, 1872 Winston Road, Charlottesville, Virginia
*Jane Garrett, Box 431, Peabody College, Nashville, Tennessee
Roman Gawkoski, 5215 North 55th Street, Milwaukee 18, Wisconsin
Richard H. Gaylord, Route 5, Box 3004, Vienna, Virginia
Helen Hofer Gee, Association of American Medical Colleges, 2530 North Ridge Avenue, Evanston, Illinois
Frederick Gehlmann, 1206 Walnut, Western Springs, Illinois
Herbert Gerjuoy, Department of Psychology, University of Toledo, Toledo 6, Ohio
Edwin E. Ghiselli, Department of Psychology, University of California, Berkeley, California
Jerome H. Gibson, 1421 Quincy, N.E., Albuquerque, New Mexico
Wilfred A. Gibson, 4009 Whann Avenue, McLean, Virginia
*David L. Giese, 106 Nicholson Hall, University of Minnesota, Minneapolis 14, Minnesota
Virginia M. (Brown) Glad, 1700 East Tufts Avenue, Englewood, Colorado
Murray Glanzer, Department of Clinical and Social Psychology, Walter Reed Army Institute of Research, Forest Glen Annex, Walter Reed Army Medical Center, Washington 12, D.C.
Robert Glaser, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
*Stuart M. Glassman, 10040 Kenberton, Oak Park 37, Illinois
Goldine C. Gleser, 3604 Lansdowne Avenue, Cincinnati 36, Ohio
Alfred F. Glixman, Department of Psychology, University of Oklahoma, Norman, Oklahoma
*Bert A. Goldman, School of Education, University of Virginia, Charlottesville, Virginia
Leon G. Goldstein, 7015 Freeport Street, Woodlawn, Hyattsville, Maryland
*Frank M. Goode, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Maxim Goode, 11912 Laurelwood Drive, Studio City, California
Leonard V. Gordon, United States Naval Personnel Research Field Activity, Building 238, United States Navy Electronic Laboratory, San Diego 52, California
Mary Agnes Gordon, 545 Club Drive, San Antonio, Texas
Mordecai H. Gordon, 4527 Chestnut Street, Bethesda 14, Maryland
David A. Grant, Department of Psychology, University of Wisconsin, Madison 6, Wisconsin
Donald L. Grant, 369 Cedar Avenue, Ridgewood, New Jersey
J. Douglas Grant, P. O. Box 365, Del Mar, California
Clifton W. Gray, c/o Department of Psychology, North Carolina State College, Raleigh, North Carolina
Bert F. Green, Jr., C-345, Box 73, Lexington 73, Massachusetts
Dorothy Green, U.S. Civil Service Commission, Room 133 Main Building, Washington 25, D.C.
Russel F. Green, 16 Riverview Heights, Rochester 23, New York

*Marshall Greenberg, Department of Psychology, University of Michigan, Ann Arbor, Michigan
William W. Grings, Department of Psychology, University of Southern California, Los Angeles 7, California
Regina C. Grudel, Psychological Service, Southbury Training School, Southbury, Connecticut
J. P. Guilford, P. O. Box 1288, Beverly Hills, California
Harold O. Gulliksen, Box 592, Princeton, New Jersey
John W. Gustad, University Counselling Service, University of Maryland, College Park, Maryland
Isaiah Guttman, Staff Research CNATECTRA, Naval Air Station, Memphis 15, Tennessee
Louis Guttman, The Hebrew University, Jerusalem, Israel
Raef K. Haddad, Bureau of Research, NJNPI, P. O. Box 1000, Princeton, New Jersey
Elizabeth Hagen, 525 West 120th Street, New York 27, New York
Ernest A. Haggard, University of Illinois College of Medicine, Department of Psychiatry, 912 South Wood Street, Chicago 12, Illinois
Clifford P. Hahn, American Institute for Research, 6135 Kansas Avenue, N.E., Washington, D.C.
Harold Wesley Hake, Department of Psychology, Gregory Hall, University of Illinois, Urbana, Illinois
Mary Hanania, Department of Mathematics, American University, Beirut, Lebanon
H. H. Harman, 520 Via de la Paz, Pacific Palisades, California
A. Edwin Harper, Jr., Ewing Christian College, Allahabad - 3, U. P., India
Mrs. Bertha P. Harper, Personnel Research Branch, TAGO, Army, Washington 25, D.C. Attn: AGTL-S
Dean H. Harper, Department of Sociology and Anthropology, University of Rochester, River Campus Station, Rochester 20, New York
Chester W. Harris, School of Education, University of Wisconsin, Madison, Wisconsin
William P. Harris, 12 Bates Road, Lexington 73, Massachusetts
Stephen Harrison, Technical Operations, Inc., 3520 Prospect Street, N.W., Washington, D.C.
Charles M. Harsh, 3110 Udall Street, San Diego 6, California
Roy Hastman, 860 Glenwood Avenue, Cincinnati 29, Ohio
Edgar M. Haverland, 5219 Danny Drive, El Paso, Texas
M. Joyce Hayes, Box 173, Dillon, South Carolina
Helen Heath, 7202 Bennett Avenue, Chicago 49, Illinois
Robert Arthur Hefner, Psychology Department, University of Michigan, Ann Arbor, Michigan
Ruth Bishop Heiser, 10 East Sharon, Glendale, Ohio
Paul Heit, Prudential Insurance Company of America, Newark 1, New Jersey
Gerald C. Helmstadter, Head, Psychometric Services, Colorado State University, Fort Collins, Colorado
E. R. Henry, 428 Demarest Avenue, Oradell, New Jersey
Sten Henrysson, Björngardsgatan 13, Stockholm So, Sweden
Miguelina N. Hernandez, Office of Evaluation, Department of Education, Hato Rey, Puerto Rico
Alfred F. Hertzka, 3411 Grand View Boulevard, Los Angeles 66, California
*Arthur H. Hill, Room 112, Psychology, University of Minnesota, Minneapolis 14, Minnesota
J. H. Hill, 22 Jump Hill Road, Levittown, Pennsylvania
*Richard E. Hilligoss, 4403 South Four Mile Run Drive, Arlington 4, Virginia

John R. Hills, Director of Testing & Guidance, Regents of the University System of Georgia, 244 Washington Street, S.W., Atlanta, Georgia
Andrew C. Hilton, 300 East 51st Street, Apt. 10-B, New York 22, New York
Elmer D. Hinckley, Box 2007, University Station, Gainesville, Florida
Michael Hirt, 1575 Lansford Drive, Columbia, South Carolina
Mrs. Danuta Hiz, 3601 Chestnut Street, Philadelphia 4, Pennsylvania
E. Lee Hoffman, Urban Life Research Institute, Tulane University, New Orleans, Louisiana
Lyle S. Hoffman, 2064 Holmby Avenue, Los Angeles 25, California
Paul J. Hoffman, Department of Psychology, University of Oregon, Eugene, Oregon
Edwin Paul Hollander, Department of Psychology, Washington University, St. Louis 5, Missouri
Peter A. Holman, 16009 Dearborn, Sepulveda, California
Wayne H. Holtzman, Department of Psychology, University of Texas, Austin 12, Texas
Paul A. Horst, Department of Psychology, University of Washington, Seattle, Washington
Alston S. Householder, Oak Ridge National Laboratory, Oak Ridge, Tennessee
William G. Howe, 267 Maplewood Drive, Rochester 15, New York
Cyril J. Hoyt, 111A Burton Hall, University of Minnesota, Minneapolis, Minnesota
William G. Hoyt, 20302 Mobile Street, Canoga Park, California
Edith M. Huddleston, 5013 Elm Street, Bethesda 14, Maryland
Lloyd G. Humphreys, Department of Psychology, University of Illinois, Urbana, Illinois
*Stephen Hunka, 1112 S. Arbor, Champaign, Illinois
Cuthbert C. Hurd, International Business Machines Corporation, 590 Madison Avenue, New York 22, New York
*John R. Hurley, Apt. 28 B, Harrison Village, Indianapolis 16, Indiana
John C. Hurst, Institute of Child Development, University of Minnesota, Minneapolis, Minnesota
Erich Hylla, Gesellschaft für Päd. Forsch. und Weiterführende Päd. Studien, E.V., Schloss-Strasse 29, Frankfurt/Main, Germany
Ray Hyman, 21 Arden Drive, Hartsdale, New York
*Hiroshi Ikeda, 845 Izumicho, Suginamiku, Tokyo, Japan
Tarow Indow, Department of Psychology, Keio University, Mita, Shiba Minatoku, Tokyo, Japan
Walter Isaac, Department of Psychology, Emory University, Atlanta 22, Georgia
Shinkuro Iwahara, Department of Psychology, Nara Women's University, Nara, Japan
Douglas N. Jackson, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania
Patricia L. Jackson, 127 East 35th Street, New York 16, New York
Thomas Owen Jacobs, USAI Human Research Unit, P. O. Box 2086, Fort Benning, Georgia
James L. Jacobson, 714 Kinnickinnic Street, Hudson, Wisconsin
Rheem F. Jarrett, 1328 Josephine Street, Berkeley 3, California
Nathan Jaspen, National League of Nursing Education, 2 Park Avenue, New York 16, New York
William E. Jaynes, Director, Bureau of Industrial Testing and Institutional Research, University of Omaha, Omaha, Nebraska
Thomas E. Jeffrey, 1013 Highland Woods, Chapel Hill, North Carolina
William A. Jenner, 7908 Kipling Parkway, S.E., Washington 28, D.C.
T. N. Jenkins, 105-21 66th Avenue, Forest Hills, New York

Thomas H. Jerdee, Industrial Relations Center, University of Minnesota, Minneapolis 14, Minnesota
Cecil D. Johnson, 406 Fenwick Drive, Falls Church, Virginia
Edward S. Johnson, Psychometric Laboratory, Nash Hall, Chapel Hill, North Carolina
M. Clemens Johnson, College of Education, University of Maryland, College Park, Maryland
Palmer O. Johnson, Department of Education, University of Minnesota, Minneapolis 14, Minnesota
Lyle V. Jones, The Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
Robert A. Jones, 1920 Clark Lane, Redondo Beach, California
Lessing A. Kahn, 12021 Milton Street, Wheaton, Maryland
Henry F. Kaiser, 1007 S. Wright Street, Champaign, Illinois
Herbert Kaizer, 1761 W. 5th Street, San Pedro, California
Richard C. Kao, 1333 Westwood Boulevard, Los Angeles 24, California
Harry Kaplan, 3705 S Street, S.E., Washington 20, D.C.
*Thomas B. Kapper, 5791 Kingsbury Place, St. Louis 12, Missouri
*George G. Karas, 24-5 Ross-Ade Drive, West Lafayette, Indiana
John E. Karlin, Bell Telephone Laboratories, Murray Hill, New Jersey
Leo Katz, Office of Naval Research Branch Office, Box 39, Navy No. 100, Fleet Post Office, New York, New York
Robert J. Keller, 130 Peik Hall, University of Minnesota, Minneapolis, Minnesota
H. Paul Kelley, Testing and Counseling Center, The University of Texas, Austin 12, Texas
T. L. Kelley, 1611 Paterna Road, Santa Barbara, California
E. Lowell Kelly, 1115 Woodlawn Street, Ann Arbor, Michigan
Maurice Kendall, The London School of Economics, Houghton Street, Aldwych, London W. C. 2, England
Norman Kettner, 8848 Lubao Avenue, Northridge, California
Mary J. Kientzle, Department of Psychology, Washington State College, Pullman, Washington
Frank Killian, Jr., USAF Hospital, Box 10208, Maxwell Air Force Base, Alabama
Herbert D. Kimmel, Route 3, Box 556K, Gainesville, Florida
Richard G. King, 10 Winslow Road, Winchester, Massachusetts
Samuel H. King, 1954 Columbia Road, N.W., Washington 9, D.C.
Rupert A. Klaus, 414 W. Clark Boulevard, Murfreesboro, Tennessee
C. James Klett, Box 653, Perry Point, Maryland
William E. Kline, 405 Terrace Way, Towson 4, Maryland
*Frederick R. Kling, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
*Eric Klinger, 5494 S. University Avenue, Chicago 15, Illinois
Henry Klugh, Alma College, Alma, Michigan
Dorothy M. Knoell, State Department of Education, 721 Capitol Avenue, Sacramento 14, California
Leopold N. H. Knops, Psychologisch Instituut, Tiense straat 96, Louvain, Belgium
Helen Lois Koch, 1374 East 57th Street, Chicago 37, Illinois
Leonard S. Kogan, Community Service Society of New York, 105 East 22nd Street, New York 10, New York
Eiichi Komiyama, 14 Nakajujo-2-chome, Kitaku, Tokyo, Japan
Samuel S. Komorita, Psychology Department, Vanderbilt University, Nashville 5, Tennessee
Robert E. Krug, Department of Psychology, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

Albert L. Kubala, Jr., Texas Women's University, Denton, Texas
Joseph F. Kubis, Fordham University, Graduate School, New York 58, New York
G. Frederic Kuder, College Station, Duke University, Durham, North Carolina
Albert K. Kurtz, Department of Psychology, University of Florida, Gainesville, Florida
David LaBerge, Psychology Department, University of Minnesota, Minneapolis, Minnesota
Oliver L. Lacey, University of Alabama, University, Alabama
William J. Laidlaw, 70 Orange Avenue, East Paterson, New Jersey
Thomas A. Lamke, Coordinator of Research, Iowa State Teachers College, Cedar Falls, Iowa
Charles R. Langmuir, The Psychological Corporation, 304 East 45th Street, New York 17, New York
Herbert Charles Lansdell, National Institute of Health, Bethesda 14, Maryland
Wilbur L. Layton, Student Counseling Bureau, 101 Eddy Hall, University of Minnesota, Minneapolis, Minnesota
Paul F. Lazarsfeld, 440 Riverside Drive, New York, New York
David B. Learner, General Motors Corporation, Box 188, North End Station, Detroit 2, Michigan
George S. Leavitt, Psychology Department, Fresno State College, Fresno 26, California
Herbert B. Leedy, 7540 East 51st Street, Lawrence, Indiana
D. Welty Lefever, 4251 Don Felipe Drive, Los Angeles 8, California
John M. Leiman, 12219 Octagon Avenue, Los Angeles 49, California
Louis T. Lepine, Veterans Administration Hospital, Coatesville, Pennsylvania
Bernard Levenson, 535 West 111th Street, New York 25, New York
Hollis M. Leverett, 21 Hampden Street, Wellesley, Massachusetts
Bernard Martin Levin, 3801 Connecticut Avenue, Washington 8, D.C.
Richard S. Levine, Educational Testing Service, Princeton, New Jersey
Don Lewis, Department of Psychology, State University of Iowa, Iowa City, Iowa
Malcolm Lichtenstein, Code 2132 Navy Electronics Laboratory, San Diego 52, California
Ernest M. Ligon, 1525 Wyoming Avenue, Schenectady 9, New York
Rikard Lindahl, Ovra Husargatan 34, Göteborg, Sweden
Richard H. Lindeman, USOM (Minn), A.P.O. No. 301, San Francisco, California
E. F. Lindquist, College of Education, State University of Iowa, Iowa City, Iowa
Gardner Lindzey, Department of Psychology, University of Minnesota, Minneapolis, Minnesota
Kenneth B. Little, Training and Standards Branch, National Institute of Mental Health, Bethesda 14, Maryland
Robert T. Littrell, Long Beach State College, 6101 East 7th Street, Long Beach 4, California
Jane Loevinger, 7209 Pershing Avenue, St. Louis 5, Missouri
Eugene Robert Long, Jr., Psychology Department, New West Building, University of North Carolina, Chapel Hill, North Carolina
Frederic Lord, 79 Dempsey Avenue, Princeton, New Jersey
Irving Lorge, Teachers College, Columbia University, New York 27, New York
Maurice Lorr, 1521 Erskine Street, Washington 12, D.C.

Ardie Lubin, Psychology Department, AMSGS, Walter Reed Army Medical Center, Washington 12, D.C.
Charles M. Lucas, 2666 South Street, Allentown, Pennsylvania
R. Duncan Luce, 4 Kirkland Place, Cambridge 38, Massachusetts
Bertha M. Luckey, Psychology Service, 1380 East 6th Street, Cleveland 14, Ohio
George A. Lundberg, Department of Sociology, University of Washington, Seattle, Washington
Erik Adrian Lundgren, Forsvarets psykologiske avdeling, Sannergaten 14, Oslo, Norway
Clifford E. Lunneborg, 5218 16th Street, N.E., Seattle, Washington
Walter A. Lurie, 742 Halstead Avenue, Mamaroneck, New York
Samuel B. Lyerly, 1204 Holly Street, Washington 12, D.C.
Howard B. Lyman, Department of Psychology, University of Cincinnati, Cincinnati 21, Ohio
R. R. Mackie, 1112 Crenshaw Boulevard, Los Angeles 19, California
Russell S. MacArthur, Faculty of Education, University of Alberta, Edmonton, Alberta, Canada
Angus G. MacLean, 5916 Hollywood Boulevard, Los Angeles 28, California
R. Daniel Malone, 346 Many Lane, Crystal Lake, Illinois
Benjamin Malzberg, Department of Mental Hygiene, State Office Building, Albany 1, New York
Gordon L. Mangan, Department of Psychology, Queen's University, Kingston, Ontario, Canada
Richard D. Mann, 112 Lexington Avenue, Cambridge, Massachusetts
Winston Howard Manning, 2700 Holiday Hill Drive, Florissant, Missouri
Melvin R. Marks, A.G.O., Personnel Research Section, Department of Army, Washington 25, D.C.
William G. Matheny, 7109 Richland Road, Fort Worth 11, Texas
*George Walter Mayeske, 502 1/2 East Stoughton, Champaign, Illinois
Sam Mayo, Loyola University, 820 Michigan Avenue North, Chicago 11, Illinois
Luzelle Denton Mays, 1534 16th Road North, Apt. 12, Arlington, Virginia
Harold L. McAdoo, 914 Barrett Road, Falls Church, Virginia
Robert McClintock, Austin Peay State College, Clarksville, Tennessee
*Ann S. McColskey, Division of Mental Health, Volusia County Health Unit, 122 North Ridgewood Avenue, Daytona Beach, Florida
Robert L. McCornack, System Development Corporation, 2500 Colorado Avenue, Santa Monica, California
William J. McGill, Department of Psychology, Columbia University, New York 27, New York
Carson McGuire, Sutton Hall 308, University of Texas, Austin, Texas
Richard McHugh, Biostatistics Division, School of Public Health, College of Medical Sciences, University of Minnesota, Minneapolis, Minnesota
*James J. McKean, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
W. B. McKean, 1105 N.E. 21st Avenue, Gainesville, Florida
K. F. Mc Laughlin, 871 N. Madison Street, Arlington 5, Virginia
Quinn McNemar, Department of Psychology, Stanford University, California
Louis L. McQuitty, Department of Psychology, Michigan State College, East Lansing, Michigan
Donald M. Medley, 535 East 80th Street, New York 21, New York
Tor Meeland, MITRE Corporation, P. O. Box 208, Lexington 73, Massachusetts
J. J. Mellinger, Personnel Research Branch, T.A.G.O., Department of Army, Washington 25, D.C.
Richard Melton, Psychology Department, Southern Illinois University, Carbondale, Illinois

Joseph R. Melville, Apt. 6B, Ridgedale Gardens, Sayre Court, Madison, New Jersey
Peter F. Merenda, 258 Negansett Avenue, Warwick, Rhode Island
Philip R. Merrifield, 629 West Wardlow Road, Long Beach 6, California
Jack C. Merwin, 200 Slocum Hall, Syracuse University, Syracuse, New York
Luigi Meschieri, Instituto Nazionale di Psycologia, Consiglio Nazionale delle Ricerce, Piazzale delle Scienze 7, Roma, Italy
Samuel J. Messick, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Charles A. Metzner, Room 2539, School of Public Health, University of Michigan, Ann Arbor, Michigan
*Milton Meux, 209 East Green Street, Champaign, Illinois
Lorenz A. Meyer, 1814 Kanawha Street, Adelphi, Maryland
William B. Michael, Director of Test Bureau, University of Southern California, University Park, Los Angeles 7, California
John E. Milholland, Psychology Department, University of Michigan, Ann Arbor, Michigan
*Arthur Lee Miller, P. O. Box 216, Chapel Hill, North Carolina
*Curtis R. Miller, 2206 South 8th Avenue, Arcadia, California
George A. Miller, Psychology Department, Memorial Hall, Harvard University, Cambridge 38, Massachusetts
*Jason Millman, 3514 Carolyn, Pittsfield Village, Ann Arbor, Michigan
*Richard Millward, Psychology Department, Indiana University, Bloomington, Indiana
John A. Modrick, 310 N. Stafford Street, Yellow Springs, Ohio
William G. Mollenkopf, 231 Hillcrest Drive, Cincinnati 15, Ohio
William T. Moran, Young and Rubicam, 285 Madison Avenue, New York, New York
Robert E. Morin, Department of Psychology, University of Texas, Austin, Texas
Donald R. Morrison, Aberdeed Drive, Lincolnshire, Crete, Illinois
Edward John Morrison, 10915 Brockbank Drive, Dallas 29, Texas
Frederick Mosteller, Department of Statistics, 2 Divinity Avenue, Cambridge 38, Massachusetts
John F. Muldoon, 200 Indian Drive, Pittsburgh 38, Pennsylvania
Carl Murchison, The Journal Press, 2 Commercial Street, Provincetown, Massachusetts
Jerome L. Myers, Department of Psychology, University of Massachusetts, Amherst, Massachusetts
Paul D. Nelson, BOQ, NAS, Corpus Christi, Texas
*Michael Nesbitt, Eno Hall, Princeton University, Princeton, New Jersey
Paul M. Neurath, 144-16 Jewel Avenue, Kew Garden Hills 67, New York, New York
J. Robert Newman, Ground Systems Division, Hughes Aircraft Company, P. O. Box 2097, Fullerton, California
*Kazuo Nihira, 1430 Beverly Boulevard, Los Angeles 26, California
Warren T. Norman, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Raymond A. Norton, Jr., Department of Psychology, George Peabody College for Teachers, Nashville 4, Tennessee
James A. Norton, Jr., Purdue University, Lafayette, Indiana
*Melvin R. Novick, 102C Isley, Chapel Hill, North Carolina
Jum C. Nunnally, Department of Psychology, University of Illinois, Urbana, Illinois
Ewald R. Nyquist, R.F.D. No. 1, Ravena, New York
Edward F. O'Day, P. O. Box B-114, San Diego State College, San Diego, California
*LeRoy A. Olson, 305 Eugenia Avenue, Madison 5, Wisconsin

Willard C. Olson, School of Education, University of Michigan, Ann Arbor, Michigan
Harry E. O'Neill, Jr., 1614 Raspberry Lane, Flint 7, Michigan
George N. Ornstein, 1544 Simpson Drive, Columbus 13, Ohio
David B. Orr, 303 Meadow Hall Drive, Rockville, Maryland
Bernice Orshansky, 1695 Grand Concourse, New York 53, New York
Hobart G. Osburn, Route No. 2, Carbondale, Illinois
*John E. Overall, 1078 Second Street, Perry Point, Maryland
Herbert Ozur, 7315 17th Avenue, West Hyattsville, Maryland
Ellis B. Page, Director of Guidance and Testing, Eastern Michigan University, Ypsilanti, Michigan
George J. Palmer, Jr., School of Business Administration, Tulane University, New Orleans, Louisiana
James W. Parker, Jr., 4505 Emerson Road, South Euclid 21, Ohio
Nancy S. Parker (Mrs. James W. Parker, Jr.), 4505 Emerson Road, South Euclid 21, Ohio
Siroon Pashalian, 189-14 Crocheron Avenue, Flushing 58, New York
Lee Edward Paul, 1743 Monticello Street, Petersburg, Virginia
Daisy Edwards Penfold, 9 Abbotsleigh Road, Streatham Park, London S. W. 16, England
Robert Perloff, 2210 W. Morse Avenue, Chicago 45, Illinois
Dallis K. Perry, 704 Manor Road, Alexandria, Virginia
Donald A. Peterson, 855 Asylum Avenue, Hartford 5, Connecticut
Robert M. Peterson, 6304 Belmar Avenue, Reseda, California
Shailer A. Peterson, American Dental Association, 222 East Superior Avenue, Chicago 11, Illinois
B. Roger Philip, 323 Annette Street, Toronto, Canada
Henri Pieron, Institut de Psychologie à la Sorbonne, Rue des Ecoles, Paris 5, France
A. E. G. Pilliner, Department of Education, 48, The Pleasance, Edinburgh 8, Scotland
Joseph A. Pittman, North Carolina College at Durham, Durham, North Carolina
Mrs. Lynnette Birdsall Plumlee, Box 63, Star Route, Cedar Crest, New Mexico
*Erich P. Prien, Jr., 2808 Avondale Road, Cleveland Heights 18, Ohio
William Prokasy, Jr., Department of Psychology, The Pennsylvania State University, University Park, Pennsylvania
*Alexander T. Quenk, 1021 Warfield Avenue, Apt. 5, Oakland 10, California
Mohammed Y. Quereshi, 1001 West Nevada Street, Urbana, Illinois
J. A. Radcliffe, Department of Psychology, Mackie Building, University of Sydney, Sydney, New South Wales, Australia
James A. Rafferty, 5008 Forest Haven Drive, Alexandria, Virginia
Mrs. Nageswari Rajaratnam, 1204 W. California, Urbana, Illinois
Robert R. Ramsey, Jr., Office of Educational Research, Box 2177, Yale Station, New Haven, Connecticut
John H. Rapparlie, R.F.D. Number 1, Box 209-B, Oak Harbor, Ohio
Gustave J. Rath, Systems and Advance Engineering, Admiral Corporation, 38 Cortland, Chicago 47, Illinois
Philburn Ratoosh, Department of Psychology, University of California, Berkeley 4, California
William S. Ray, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania
James C. Reed, Department of Reading & Study Skills, Wayne State University, Detroit, Michigan
Olav Reiersøl, Matematisk Institut, Universitetet, Bindern, Oslo, Norway
James J. Reilly, 352 Orange Street, Manchester, New Hampshire

Wm. Reitz, 727 Student Center, Wayne State University, Detroit 2, Michigan
Frank Restle, c/o Department of Psychology, Michigan State University, East Lansing, Michigan
Mary E. Reuder, 144 - 70C Village Road, Parkway Village, Jamaica 35, New York
Jack Richardson, Department of Psychology, Harpur College, Endicott, New York
Charles V. Riche, Jr., 12 Orchard Road, Bedford, Massachusetts
James H. Ricks, Jr., The Psychological Corporation, 304 East 45th Street, New York 7, New York
Marilyn K. Rigby, 3680 Lindell Boulevard, St. Louis 8, Missouri
Joseph W. Rigney, 7303 Cleargrove Drive, Downey, California
Horacio J. A. Rimoldi, 5640 Kimbark Avenue, Chicago 37, Illinois
Darrell Eugene Roach, Nationwide Insurance Companies, 246 North High Street, Columbus, Ohio
Marguerite E. Robertson, 221 Adams Street, Warwick, Rhode Island
W. S. Robinson, Department of Anthropology and Sociology, University of California, Los Angeles 24, California
William L. Robinson, Jr., 1512 A Brackenridge Apts., Austin, Texas
Pablo Roca, P. O. Box 264, Hato Rey, Puerto Rico
Milton L. Rock, 121 S. Broad Street, Philadelphia 7, Pennsylvania
Merrill Roff, University of Minnesota, Institute of Child Welfare, Minneapolis 14, Minnesota
Miles S. Rogers, 9412 Isis Avenue, Los Angeles 45, California
Robert C. Rogers, R.F.D. No. 1, Princeton, New Jersey
Albert C. Rohloff, 476 Roff Avenue, Palisades Park, New Jersey
Nathan Rosenberg, 2800 Quebec Street, N.W., Washington 8, D.C.
Benjamin Rosner, Assistant Professor of Education, Brooklyn College, Bedford Avenue and Avenue H, Brooklyn, New York
*Donald Clare Ross, Nash Hall, Chapel Hill, North Carolina
Paul F. Ross, Employee Relations Department, Standard Oil Company of New Jersey, 30 Rockefeller Plaza, New York 20, New York
Yvonne A. Rubens, 116-44 169 Street, Jamaica 34, New York
Stanley Arthur Rudin, 102 Greenhill Pass, San Antonio 1, Texas
*Britton K. Ruebush, 21 Elmwood Road, New Haven, Connecticut
Phillip J. Rulon, Peabody House, 13 Kirkland Street, Cambridge 38, Massachusetts
J. Francis Rummel, School of Education, University of Oregon, Eugene, Oregon
Edward A. Rundquist, 1405 Summit Avenue, Alexandria, Virginia
Philip J. Runkel, Bureau of Educational Research, University of Illinois, 1007 S. Wright Street, Champaign, Illinois
Jean M. Rutherford, Northern Michigan College, Marquette, Michigan
T. A. Ryan, Department of Psychology, Morrill Hall, Cornell University, Ithaca, New York
Robert Sadacca, 8110 Tahona Drive, Silver Spring, Maryland
Edward Sait, U. S. Army Enlisted Evaluation Center, Fort Benjamin Harrison, Indianapolis, Indiana
James M. Sakoda, Department of Psychology, University of Connecticut, Storrs, Connecticut
Richard T. Sampson, Nationwide Insurance, 246 N. High Street, Columbus 16, Ohio
Bryan B. Sargent, III, 1620 Highland Avenue, Knoxville, Tennessee
David R. Saunders, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Jack Sawyer, Department of Statistics, University of Chicago, Chicago 37, Illinois

Douglas E. Scates, University of Florida, Gainesville, Florida
K. Warner Schaie, Department of Psychology, University of Nebraska, Lincoln 8, Nebraska
Ivan Scheier, 203 East Green Street, Apartment 1, Champaign, Illinois
Kenneth F. Schenkel, Training Department (89-01), Lockheed Aircraft Corporation, Marietta, Georgia
*Harold Schiffman, Eno Hall, Princeton University, Princeton, New Jersey
Richard A. Schindler, 2109 C Cedar Street, Berkeley 9, California
Lowell Schipper, Department of Psychology, Kansas State College, Manhattan, Kansas
Morris Schnore, Psychology Department, University College, University of Western Ontario, London, Ontario, Canada
William B. Schrader, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Douglas Schultz, 122 West Suburban Avenue, State College, Pennsylvania
Charles F. Schumacher, 2530 Ridge Avenue, Evanston, Illinois
Richard E. Schutz, Arizona State University, Tempe, Arizona
Paul A. Schwarz, American Institute for Research, 410 Amberson Avenue, Pittsburgh 32, Pennsylvania
William C. Schwarzbek, Public and Employee Relations Services, General Electric Company, 570 Lexington Avenue, New York 22, New York
Robert Schweiker, 23 Wendell Street, Cambridge, Massachusetts
Robert W. Scollay, Leroy N. Vernon and Associates, 221 North LaSalle Street, Chicago 1, Illinois
Lee B. Sechrest, Department of Psychology, Northwestern University, Evanston, Illinois
Dean W. Seibel, 19 Palmer Road, Franklin Park, New Jersey
Robert Seibel, 807 King Street, Peekskill, New York
Saul B. Sells, 3204 Sweetbrier Lane, Fort Worth 9, Texas
Richard B. Seymour, 1614 W. 213th Street, Torrance, California
Marion F. Shaycoft, 7620 Maple Avenue, Apartment 407, Washington, D.C.
Maynard W. Shelly, 843 Holmes Avenue, Deerfield, Illinois
Roger N. Shepard, Department 122, Bell Telephone Laboratories, Murray Hill, New Jersey
Joseph Russell Sherlock, 2685 Briggs Avenue, New York 58, New York
R. H. Shevenell, Institute of Psychology, University of Ottawa, Ottawa, Canada
Nathan Shock, Baltimore City Hospitals, Unit on Gerontology, Baltimore 24, Maryland
Emir H. Shuford, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
Meyer Shultz, 1812 Allison Street, N.W., Washington, D.C.
Herbert S. Sichel, Operational Research Bureau, P. O. Box 10238, Johannesburg, South Africa
Paul S. Siegel, Department of Psychology, University of Alabama, University, Alabama
Sidney Siegel, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania
George B. Simon, Lauderdale Road, Mail Route 98, Binghamton, New York
Roy G. Simpson, 937 Avenue C, Fort Madison, Iowa
Rosedith Sitgreaves, Box 126, Teachers College, Columbia University, New York 27, New York
*Malcolm J. Slakter, Mathematics Department, Smith, Syracuse University, Syracuse 10, New York
J. E. Keith Smith, 14 Audubon Road, Wellesley, Massachusetts
Max Smith, 453 E. 96th Street, Brooklyn 12, New York

*Timothy A. Smith, Box 466, Chapel Hill, North Carolina
*Marvin Snider, Apt. 77, 4 O'Leary Way, Jamaica Plain 30, Massachusetts
Herbert Solomon, Statistics Department, Stanford University, Stanford, California
William D. Spears, Department of Foundations, School of Education, University of Florida, Gainesville, Florida
*Douglas K. Spiegel, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
Richard E. Stafford, 216-B Halsey Street, Princeton, New Jersey
Robert Stake, 3347 Starr, Lincoln, Nebraska
John M. Stalnaker, 569 Briar Lane, Northfield, Illinois
Julian C. Stanley, Jr., Education Building, University of Wisconsin, Madison 6, Wisconsin
Artie John Stauffer, 628 Island View Drive, Seal Beach, California
Leonard W. Staugas, 3214 Pearl Street, Santa Monica, California
John E. Stecklein, 211 Burton Hall, University of Minnesota, Minneapolis, Minnesota
W. R. Stellwegen, School of Education, University of California, Berkeley 4, California
Frederick F. Stephan, 105 FitzRandolph Road, Princeton, New Jersey
Theodor D. Sterling, Director, Biometric Section, Division Preventive Medicine, College of Medicine, Eden and Bethesda Avenue, Cincinnati, Ohio
Jack J. Sternberg, 5118-72 Place, Glenridge, Hyattsville, Maryland
*Saul H. Sternberg, 46 Irving Street, Cambridge 38, Massachusetts
E. Elizabeth Stewart, 17 Maxwell Road, Glen Lennox, Chapel Hill, North Carolina
Naomi Stewart, No. 4, Wick Hall, Abingdon, Berkshire, England
Glen F. Stice, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
Donald W. Stilson, Department of Psychology, University of Colorado, Boulder, Colorado
Pairlee J. Stinson, 7651 Carrswold Drive, Clayton 5, Missouri
James H. Straughan, 1290 Garden Street, Palo Alto, California
Elmer Louis Streuning, 545 West 111th Street, Apt. 6L, New York, New York
Patrick Suppes, Serra House, Stanford University, Stanford, California
Frances Swineford, Educational Testing Service, Box 592, Princeton, New Jersey
Arthur T. Tait, 5523 Rumsey Drive, Riverside, California
Alfonso Tallet-Grenier, Universidad de Oriente, Santiago de Cuba, Cuba
*Jack T. Tapp, 1008 S. Second Street, Champaign, Illinois
Merle W. Tate, 3812 Walnut Street, Philadelphia 4, Pennsylvania
Calvin W. Taylor, Department of Psychology, University of Utah, Salt Lake City, Utah
Erwin K. Taylor, 314 Superior Avenue, Cleveland 14, Ohio
Harold C. Taylor, Director, W. E. Upjohn Institute for Community Research, 709 South Westnedge Avenue, Kalamazoo 4, Michigan
Howard Tempero, 309 A Teachers College, University of Nebraska, Lincoln, Nebraska
*Donald B. Thomas, Department of Psychology, Princeton University, Princeton, New Jersey
J. G. Thomas 95, 610 West 115th Street, New York 25, New York
James A. Thomas, 233 Oglethorpe Street, N.E., Washington 11, D.C.
Kenneth F. Thomson, 61 Highfield Avenue, Port Washington, New York
Robert L. Thorndike, Teachers College, Columbia University, New York 27, New York

Robert M. Thrall, Department of Mathematics, University of Michigan, Ann Arbor, Michigan
Thelma G. Thurstone, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina
David V. Tiedeman, Peabody House, 13 Kirkland Street, Cambridge 38, Massachusetts
Kalman Tillem, 10710 Corning, Oak Park, Michigan
Helen Tomlinson, 354 Broadview West, San Antonio, Texas
Herbert A. Toops, 1430 Cambridge Boulevard, Columbus 8, Ohio
Warren S. Torgerson, M.I.T., Lincoln Laboratory, Lexington 73, Massachusetts
Leonard W. Towner, Department of Psychology, Long Beach State College, Long Beach, California
Arthur E. Traxler, 21 Audubon Avenue, New York 32, New York
David K. Trites, Personnel Laboratory, WADC, Box 1557, Lackland Air Force Base, Texas Attn: WCLLR
Theresa G. Trittipoe, American Institute for Research, 6135 Kansas Avenue, N.E., Washington 11, D.C.
William F. Tronca, 2506 30th South, Seattle 44, Washington
*William Trousdale, 4925 Fourth Street, N.E., Minneapolis, Minnesota
Robert C. Tryon, Department of Psychology, University of California, Berkeley 4, California
Ledyard R Tucker, 102 Cedar Lane, Princeton, New Jersey
Read D. Tuddenham, Department of Psychology, University of California, Berkeley 4, California
*Mehmet F. Turgut, 517 East 138th Street, Bronx 54, New York
William Watson Turnbull, 138 Westcott Road, Princeton, New Jersey
Dik W. Twedt, 510 Linden Avenue, Wilmette, Illinois
Fred T. Tyler, Haviland Hall, University of California, Berkeley, California
Ralph W. Tyler, Center for Advanced Study in Behavioral Sciences, Inc., 202 Junipero Serra, Stanford, California
E. J. Uhlaner, 5430 N. 20th Street, Arlington 5, Virginia
C. C. Upshall, 21 Eastview Terrace, Pittsford, New York
Harry S. Upshaw, School of Public Health, University of North Carolina, Chapel Hill, North Carolina
Steven G. Vandenberg, Mental Health Research Institute, University of Michigan, Ann Arbor, Michigan
James M. Vanderplas, Psychology Department, Washington University, St. Louis, Missouri
John van Laer, Department of Psychology, Northwestern University, Evanston, Illinois
Neil J. Van Steenberg, 5556 North 16th Street, Arlington, Virginia
Marvin J. Van Wagenen, 1729 Irving Avenue South, Minneapolis 5, Minnesota
Manmohan Varma, Professor of Education, Gorakhpur University, Gorakhpur, U.P., India
*Jan Vastenhouw, 11 Kamperfoeliestraat, Bussum, Netherlands
C. L. Vaughn, Psychological Corporation, 304 East 45th Street, New York 17, New York
Helen M. Walker, Teachers College, Columbia University, New York 27, New York
William B. Walker, General Electric Company, Public Relations Research Services, 570 Lexington Avenue, New York, New York
David L. Wallace, Department of Statistics, University of Chicago, Chicago 37, Illinois
Wimburn L. Wallace, The Psychological Corporation, 304 East 45th Street, New York 17, New York

Charles L. Walter, Psychology Service, Veterans Administration Hospital, Salisbury, North Carolina
M. J. Wantman, Testing and Counseling Service, Morey Hall, Room 204, University of Rochester, Rochester 20, New York
Joe H. Ward, Jr., 167 East Arrowhead Drive, San Antonio 1, Texas
Edward E. Ware, Institute of Communications Research, University of Illinois, Urbana, Illinois
Neil D. Warren, 3827 Olympiad Drive, Los Angeles 43, California
Lawrence K. Waters, U. S. Naval School of Aviation Medicine, Naval Air Station, Pensacola, Florida
Richard W. Watkins, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
*Francis R. Watson, 400 E. Northwood, Columbus 1, Ohio
Herbert B. Weaver, Department of Psychology, University of Hawaii, Honolulu 14, Hawaii
Sam C. Webb, Department of Psychology, Emory University, Atlanta 22, Georgia
Harold Webster, Center for the Study of Higher Education, 2747 Bancroft Way, Berkeley 4, California
*Murray Weiner, 137-76 70th Avenue, Kew Gardens Hills 67, New York
Arla Weinert, 905 Fairview Avenue, Apartment 3, Washington 12, D.C.
Frederic D. Weinfeld, Graduate School of Education, Harvard University, Cambridge 38, Massachusetts
Eleanor S. Weiss, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey
*Ronald A. Weitzman, Eno Hall, Princeton University, Princeton, New Jersey
Eugene H. Welsand, 3411 S. 10th Street, Milwaukee 15, Wisconsin
George Welsh, Psychology Department, University of North Carolina, Chapel Hill, North Carolina
Alexander G. Wesman, 84 Ralph Avenue, White Plains, New York
Risdon John Westen, Evaluation Directorate, Office of the Registrar, U.S. Air Force Academy, Colorado
*Leonard Wevrick, Department of Psychology, University of Sydney, Sydney, New South Wales, Australia
Robert J. Wherry, Department of Psychology, Ohio State University, Columbus 10, Ohio
Milton Whitcomb, Psychology Department, University of Texas, Austin, Texas
Benjamin W. White, Lincoln Laboratory, M.I.T., Box 73, Lexington 73, Massachusetts
Dean K. Whitla, Office of Tests, 11 University Hall, Cambridge 38, Massachusetts
Alfred G. Whitney, Life Insurance Agency Management Association, 855 Asylum Avenue, Hartford 5, Connecticut
S. S. Wilks, Department of Mathematics, Princeton University, Princeton, New Jersey
Richard W. Willard, P. O. Box 105, Littleton, Massachusetts
Louis P. Willemin, 11028 Burnley Terrace, Silver Spring, Maryland
*Bob J. Williams, Department of Psychology, University of Illinois, Urbana, Illinois
*Paul L. Williams, 307 E. Higgins Road, Elk Grove Village, Illinois
Robert J. Williams, 142 Linden Avenue, Emerson, New Jersey
Warren W. Willingham, World Book Company, Yonkers, New York
Norman E. Willmorth, 20632 Haynes, Canoga Park, California
Clark L. Wilson, 1010 E. Summit Drive, Whittier, California
Kellogg V. Wilson, Department of Psychology, Kent State University, Kent, Ohio

Ben James Winer, Department of Psychology, Purdue University, Lafayette, Indiana
Herman O. A. Wold, University Institute of Statistics, Järnbrogatan 18, Uppsala, Sweden
Morris Wolfe, 1215 S. 4th Street, Champaign, Illinois
Dael L. Wolfle, 2801 Rittenhouse Street, N.W., Washington 15, D.C.
William David Wolking, 4831 28th Avenue South, Minneapolis 17, Minnesota
Ben D. Wood, Columbia University, New York 27, New York
Max A. Woodbury, 401 West 205th Street, New York 34, New York
Calvin E. Wright, 2918 Fuhrman Avenue, Seattle 2, Washington
J. Wayne Wrightstone, 21 Hickory Road, Summit, New Jersey
C. F. Wrigley, Department of Psychology, Michigan State University, East Lansing, Michigan
Matsusaburo Yokoyama, Department of Psychology, Keio University, Mita, Tokyo, Japan
Francis A. Young, Department of Psychology, Washington State College, Pullman, Washington
Irving Yudelowitz, 11 Midwood Street, Brooklyn 25, New York
Virginia Zachert, P. O. Box 671, Auburn, Alabama
Joseph Zeidner, 1001 South Belgrade Road, Silver Spring, Maryland
Robert C. Ziller, Group Dynamics Center, University of Delaware, Newark, Delaware
Joseph L. Zinnes, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Joseph Zubin, 722 W. 168th Street, New York, New York

Minutes of the
1959 ANNUAL BUSINESS MEETING
of the
PSYCHOMETRIC SOCIETY

The regular Annual Meeting of the Psychometric Society was held in Cincinnati, Ohio, on Tuesday, September 8, 1959. President Frederic M. Lord called the meeting to order at 2:00 P.M.

The minutes of the previous Annual Meeting were approved.

On a ballot for the election of two new members of the Council of Directors, Dr. Jane Loevinger and Dr. John E. Milholland were elected for a term of three years, ending in 1962.

Dr. Lloyd G. Humphreys reported for the Membership Committee. The Membership Committee nominated 45 persons as full members, 60 student members to be transferred to full membership, and 32 individuals as student members.

It was moved, seconded, and passed that the following 45 persons be elected as full members:

Frank B. Baker
Donald B. Black
Lyle E. Bourne
Allen C. Busch
Don J. Cosgrove
John W. Cotton
William F. Dossett
S. David Farr
David J. Fitch
Raymond W. Frankman
Paul A. Games
Clifford P. Hahn
Dean Harper
Paul Heit
Miguelina N. Hernandez
Edwin P. Hollander
John G. Hurst
Thomas Owen Jacobs
John E. S. de Jung
William J. Laidlaw
Herbert C. Lansdell
Russell S. MacArthur
R. Daniel Malone
Sam Mayo
Jum C. Nunnally
Ellis B. Page
George J. Palmer, Jr.
Mohammed Y. Quereshi
J. A. Radcliffe
Frank Restle
Lillian Capo de Rivero
W. S. Robinson
Nathan Rosenberg
Richard A. Schindler
Charles F. Schumacher
Elmer Louis Streuning
Theresa G. Trittipoe
Read D. Tuddenham
David L. Wallace
Wimburn Wallace
Sam C. Webb
Frederic D. Weinfeld
Warren W. Willingham
Francis A. Young
Joseph Zeidner

It was moved, seconded, and passed that the 60 student members listed below be transferred to full membership:

Nancy S. Anderson
Norman Anderson
David C. Beardslee
Morton A. Bertin
Nicholas A. Bond, Jr.
Eugene A. Bouvier
James G. Boyce
Thomas H. Burkhalter
Ralph McC. Chinn
Paul Raymond Christensen
Norman Cliff
Bernard P. Cohen
W. C. Coppock
Mary Corcoran

Lolafaye Coyne
Elliott M. Cramer
Kern W. Dickman
Sidney Epstein
Donald Estavan
Gordon Fifer
William J. Flynn, Jr.
Roman Gawkoski
Clifton W. Gray
Isaiah Guttman
Roy Hastman
Robert Arthur Hehner
Shinkuro Iwahara
James L. Jacobson
Thomas H. Jerdie
Henry F. Kaiser
Herbert D. Kimmel
Rupert A. Klaus
Henry E. Klugh
Richard H. Lindemen
Clifford E. Lunneborg
Richard D. Mann
Robert E. McClintock
John F. Muldoon
Warren T. Norman
David B. Orr
Lee Edward Paul
William Prokasy, Jr.
James C. Reed
Charles V. Riche, Jr.
William L. Robinson, Jr.
Robert Sadacca
K. Warner Schaie
Morris Schnore
Richard E. Schutz
Robert W. Scollay
Ahmed El Sensussi
Roy G. Simpson
Richard E. Stafford
Robert Earl Stake
E. Elizabeth Stewart
J. G. Thomas
Edward E. Ware
Richard W. Willard
Calvin E. Wright
Joseph L. Zinnes

It was moved, seconded, and passed that the 32 persons named below be elected as student members.

G. Ernest Anderson, Jr.
Joan A. Barr
Alan R. Bass
Carl E. Bereiter
Robert Besco
Vincent N. Campbell
Ronald Lynn Flaugher
Jane Garrett
Bert A. Goldman
Marshall Greenberg
David L. Grese
Richard Earl Hilligoss
Stephen M. Hunka
Edward S. Johnson
Francis B. Kapper
Eric Klinger
George Walter Mayeske
James J. McKeon
Arthur Lee Miller
Curtis R. Miller
Jason Millman
Richard Millward
Alexander T. Quenka
Donald Clare Ross
Malcolm J. Slakter
Timothy A. Smith
Saul H. Sternberg
Jack T. Tapp.
Mehmet F. Turgut
Bob J. Williams
Paul L. Williams
Morris Wolfe

It was moved, seconded, and passed that the Membership Committee be thanked for their excellent work.

Dr. Wilfred A. Gibson reported for the Program Committee. Of twelve abstracts of papers submitted for consideration for presentation at the Annual Meeting, ten were accepted. One symposium was scheduled. The Committee cooperated with the Division 5 Program Committee in finding a speaker for the joint dinner and in scheduling the various technical and business sections. It was moved and seconded that the report be accepted with thanks. Motion passed.

Dr. Harold Gulliksen reported for the Committee on the 25th Anniversary of the Psychometric Society. He reported general agreement upon the importance of an appropriate celebration. It was moved and seconded that the present Committee for the 25th Anniversary be discharged with thanks and that the incoming president appoint a committee of three to implement an appropriate plan for the occasion. Motion passed.

It was moved and seconded that the Society appropriate a sum of up to $500.00 to be expended in connection with the 25th Anniversary celebration of the Society. Motion passed.

It was moved and seconded that, in connection with the annual billings for dues, the Treasurer invite voluntary contributions from the membership for executing plans in connection with the 25th Anniversary celebration, with $2.00 suggested as a typical contribution. Motion passed.

Dr. Charles F. Wrigley pointed out that 1960 is not only the 25th Anniversary of the Psychometric Society, but also the 100th anniversary of Fechner's book and that it might be appropriate for the Psychometric Society to pay attention to this important anniversary. It was moved and seconded that this fact be brought to the attention of the committee arranging plans for the 25th Anniversary, together with the possibility of collaborating with the American Psychological Association in the matter of the Fechner centennial. Motion passed.

The report of the Treasurer was presented by Dr. William B. Schrader. A copy is attached. The report was accepted with thanks.

Dr. Ledyard R Tucker reported for the Auditing Committee. The Treasurer's books, bills received, cancelled checks, and bank statements for the fiscal year July 1, 1957 to June 31, 1958 were inspected. All financial matters of the Society were found to be in good form and in accord with the report published in the December 1958 issue of Psychometrika. The report was accepted with thanks.

On ballot, Dr. Philip H. DuBois was re-elected Secretary of the Psychometric Society for a term of three years, ending September 30, 1962.

Dr. Wrigley reported for the Committee on Special Memberships. The Committee suggested that a category of "Foreign Affiliates" be established. Foreign affiliates would receive Psychometrika, would have the privilege of participating in the meetings of the Society, and would pay the same dues as student members; however, no endorsers would be required on applications and foreign affiliates would not have the right of holding office. Persons residing abroad who wish to become regular members could continue to do so, as at the present time. It was moved and seconded that steps be taken to establish this category of membership. Motion passed.

The Committee on Special Memberships also suggested that there be established a category of "Emeritus Members" for individuals who have reached the age of 65 and who have paid full dues for 15 years. Such emeritus members would receive Psychometrika and would have full membership privileges, including the right to vote and to hold office. Suggested dues would be $1.00 a year. It was moved and seconded that steps be taken to establish such a category. Motion passed.

The Secretary's report was presented by Dr. DuBois. It was accepted with thanks.

Reporting for the Election Committee, Dr. Clyde H. Coombs stated that Dr. Humphreys has been elected President of the Psychometric Society for a period of one year, beginning October 1, 1959.

The meeting was adjourned at 5:00 P. M.

Philip H. DuBois
Secretary

PSYCHOMETRIC SOCIETY

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1959

RECEIPTS (Dues)

Year	Members	Student Members
1959	585	51
1958	40	8
1957	12	1
1956	4	-
	641	60

	$4,727.00
Received with Dues for Corporation Publications	227.56
Net overpayments	1.35
Total Receipts	$4,955.91

DISBURSEMENTS

Psychometric Corporation (90% of dues)	$4,254.30
Psychometric Corporation (Publications)	227.56
Stationery and Postage	385.90
Secretarial Services	214.88
Bank Charges	4.09
Telephone	16.18
Refund	1.00
Total Disbursements	$5,103.91

BALANCE

Balance, June 30, 1958	$1,345.20
Receipts, 1958-59	4,955.91
	$6,301.11
Disbursements, 1958-59	5,103.91
Balance, June 30, 1959	$1,197.20

PSYCHOMETRIC CORPORATION

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1959

RECEIPTS

Subscriptions (less agency discounts)	$6,799.80
Psychometric Society (90% of dues)	4,254.30
Sale of Back Issues (less discounts)	2,490.71
Sale of Monographs 5-8 (less discounts)	145.30
Interest on Savings Accounts	275.63
Reprints	762.15
Net overpayments	2.24
Mailing List Use	5.07
	$14,735.20

DISBURSEMENTS

Printing and Mailing *Psychometrika* Volume 23, No. 2, through 24, No. 1	$7,696.40
Reprints	477.42
Stipend of Managing Editor (7/1/58 - 6/30/59)	750.00
Stipend of Assistant Editor (7/1/58 - 6/30/59)	500.00
Stipend of Treasurer (7/1/58 - 6/30/59)	250.00
Secretarial Services: Editorial Office	750.00
Secretarial Services: Business Office	296.74
Stationery and Postage	218.88
Mailing Back Issues and Monographs	156.87
Psychometric Society (Incorrect credit)	28.00
Refund	1.00
Legal Fees	35.00
Mailing List Use	5.07
Miscellaneous	27.00
	$11,192.38

BALANCE AND RESERVES

Balance, June 30, 1958	$7,242.74
Reserve Funds, June 30, 1958	
Englewood Savings and Loan Assn. Englewood, Colorado	3,500.00
Metropolitan Savings and Loan Assn. Los Angeles, California	3,500.00
Total	$14,242.74
Receipts, 1958-59	14,735.20
Sum	28,977.94
Disbursements	11,192.38
Remainder	$17,785.56
Balance, June 30, 1959	$10,785.56
Reserve Funds, June 30, 1959	
Englewood Savings and Loan Assn. Englewood, Colorado	3,500.00
Metropolitan Savings and Loan Assn. Los Angeles, California	3,500.00
Total, Balance and Reserve Funds	$17,785.56

OBLIGATIONS

Estimated cost of *Psychometrika*, Vol. 24, Nos. 2-4 Printing and Mailing	$6,000.00
Stipends (7/1/59 - 12/31/59)	750.00
Secretarial Services	550.00
	$7,300.00

BALANCE AND RESERVES, LESS OBLIGATIONS	$10,485.56

ERRATUM

In Cureton, Edward E., Note on $\phi/\phi_{\max}$. *Psychometrika*, 1959, 24, 89–92, the first sentence of paragraph 2, page 89, should read "It is well known that ϕ can equal $+1$ only if $p_1 = p_2$, and that it can equal -1 only if $p_1 = p_2$ ([1], p. 324; [2], p. 342)."

INDEX FOR VOLUME 24

Bock, R. Darrell. Review of "P. Oléron, *Les Composantes de l'Intelligence d'après les Recherches Factorielles.*" 187-188.

Bryant, Steven J. (with J. G. Marica). Strategies and learning models. 253-256.

Colver, Robert M. Estimating item indices by nomographs. 179-185.

Cureton, Edward E. Note on $\phi/\phi_{\max}$. 89-91.

Cureton, Edward E. A note on factor analysis: arbitrary orthogonal transformations. 169-174.

DuBois, Philip H. Review of "*L'Analyse Factorielle et ses Applications.*" 376-378.

Edwards, Allen L. A note on Tryon's measure of reliability. 257-260.

Elfving, G. (with R. Sitgreaves and H. Solomon). Item selection procedures for item variables with a known factor structure. 189-205.

Fagot, Robert F. A model for ordered metric scaling by comparison of intervals. 157-168.

Garner, Wendell R. Review of "W. S. Torgerson, *Theory and Methods of Scaling.*" 277-279.

Geisser, Seymour (with S. W. Greenhouse). On methods in the analysis of profile data. 95-112.

Gibson, W. A. Three multivariate models: factor analysis, latent structure analysis, and latent profile analysis. 229-252.

Goodman, Leo A. Simple statistical methods for scalogram analysis. 29-43.

Greenhouse, Samuel W. (with S. Geisser). On methods in the analysis of profile data. 95-112.

Hanania, Mary I. A generalization of the Bush-Mosteller model with some significance tests. 53-68.

Hoffman, Paul J. Generating variables with arbitrary properties. 265-267.

Kaiser, Henry F. A note on the Tryon-Kaiser solution for the communalities. 269-271.

Katz, Leo (with C. H. Proctor). The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process. 317-327.

Kish, Leslie. Review of "P. J. McCarthy, *Introduction to Statistical Reasoning.*" 280-281.

Kline, William E. A synthesis of two factor analyses of intermediate algebra. 343-359.

Lord, Frederic M. Statistical inferences about true scores. 1-17.

Lord, Frederic M. Randomly parallel tests and Lyerly's basic assumption for the Kuder-Richardson formula (21). 175-177.

Lord, Frederic M. An approach to mental test theory. 283-302.

McConnell, David. An augmented model for spontaneous regression and recovery. 145-155.

Madansky, Albert. Least squares estimation in finite Markov processes. 137-144.

Marica, John G. (with S. J. Bryant). Strategies and learning models. 253-256.

Maxwell, A. E. Maximum likelihood estimates of item parameters using the logistic function. 221-227.

Messick, Samuel. Review of "C. Coombs and R. C. Kao, *Nonmetric Factor Analysis.*" 93-94.

Proctor, Charles H. (with L. Katz). The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process. 317-327.

Prokhovnik, S. J. Pattern variants on a square field. 329-341.

Restle, Frank. A metric and an ordering on sets. 207-220.

Sagi, Philip C. A statistical test for the significance of a coefficient of reproducibility. 19-27.

Savage, I. Richard. Review of "D. A. S. Fraser, *Nonparametric Methods in Statistics.*" 279-280.

Schucker, R. E. A note on the use of triads for paired comparisons. 273-276.

Schutz, William C. Some implications of the logical calculus for empirical classes for social science methodology. 69-87.

Siegel, Sidney. Theoretical models of choice and strategy behavior: stable state behavior in the two-choice uncertain outcome situation. 303-316.

Sitgreaves, R. (with G. Elfving and H. Solomon). Item selection procedures for item variables with a known factor structure. 189-205.
Smoke, William H. (with R. B. Zajonc). Redundancy in task assignments and group performance. 361-369.
Solomon, H. (with G. Elfving and R. Sitgreaves). Item selection procedures for item variables with a known factor structure. 189-205.
Thrall, R. M. Review of "D. Davidson, P. Suppes, and S. Siegel, *Decision Making, An Experimental Approach.*" 277.
Travers, Robert M. W. Review of "J. Loevinger, *Objective Tests as Instruments of Psychological Theory.*" 375-376.
Tryon, Robert C. Domain sampling formulation of cluster and factor analysis. 113-135.
Wherry, Robert J. Hierarchical factor solutions without rotation. 45-51.
Weiner, Irving B. A note on the use of Mood's likelihood ratio test for item analyses involving 2 × 2 tables with small samples. 371-372.
Wolins, Leroy. An improved procedure for the Wherry-Winer method for factoring large numbers of items. 261-264.
Zajonc, Robert B. (with W. H. Smoke). Redundancy in task assignments and group performance. 361-369.

BUREAU OF EDNL. & PSYCL. RESEARCH